Development of a PTSD terminology framework to support knowledge acquisition, concept recognition, and data mining

by

Bryan Gamble

A DISSERTATION

Presented to the Department of Medical Informatics and Clinical Epidemiology
and the Oregon Health & Science University
School of Medicine
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

June 2016

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the PhD Dissertation of

B. Travis Gamble

"Development of a PTSD terminology framework to support knowledge acquisition, concept recognition, and data mining"

Has been approved

_____

Aaron Cohen, MD

_____

Melissa Haendel, PhD

_____

Steven Bedrick, PhD

_____

Thomas Rindflesch, PhD

_____

Blake Lesselroth, MD

_____

Jianji Yang, PhD

TABLE OF CONTENTS

TABLE OF CONTENTS i
LIST OF FIGURES iv
LIST OF TABLES vi
ACKNOWLEDGEMENT vii
ABSTRACT ix
PREFACE xi

1 INTRODUCTION 1
1.1 Aggregated Terminology for PTSD 1
1.2 Problem Statement 3
1.3 Significance and Motivation 4
1.4 Primary Research Question 4
1.5 Overview of Research Aims and Hypotheses 6
1.5.1 Development of a Terminology System (Aim 1) 6
1.5.2 Text Mining Pipeline Implementation (Aim 2) 8
1.5.3 Data Mining Analyses (Aim 3) 9
1.6 Contribution 11
1.7 Dissertation Structure 12

2 BACKGROUND 13
2.1 Outline 13
2.2 Post-traumatic Stress Disorder (PTSD) 13
2.2.1 PTSD Prevalence 14
2.2.2 PTSD Terminology Initiatives 15
2.3 Terminology Structures and Formats 17
2.4 Medical Terminology Resources 20
2.5 Mental Health Terminology Resources 24
2.6 Terminology Development 26
2.7 Biomedical Text Mining 28
2.8 Natural Language Processing (NLP) Frameworks 31
2.8.1 Selected NLP Frameworks 31
2.8.2 Selected Text Processing Systems Pipelines 33
2.9 Data Mining 36
2.9.1 Selected Data Mining Algorithms 36
2.9.2 Selected Machine Learning Toolkits 39
2.10 Gaps in Literature and Knowledge 40

3 PTSD TERMINOLOGY SYSTEM 45
3.1 Introduction 45
3.2 PTSD Terminology Development 47
3.2.1 Post-traumatic Stress Disorder (PTSD) 47
3.2.2 Terminology Systems 47
3.2.3 Needs Assessment 49

3.2.4 Requirements Engineering 50
3.2.5 Knowledge Acquisition 51
3.2.6 Concept Overlap 52
3.3 Description of Aim 1 53
3.3.1 Conceptual Model 53
3.3.2 Research Questions and Hypotheses 54
3.4 Methods 55
3.4.1 Needs Assessment 56
3.4.2 Requirements Engineering 59
3.4.3 Terminology System Development 65
3.4.4 Knowledge Acquisition 69
3.4.5 Concept and Term Overlap 73
3.5 Results 74
3.5.1 Needs Assessment 74
3.5.2 Terminology Requirements 77
3.5.3 PTSD Terminology System Development 79
3.5.4 Knowledge Acquisition Saturation 80
3.5.5 Concept and Term Overlap 86
3.6 Discussion 90
3.7 Conclusion 94

4 PTSD TEXT MINING 96
4.1 Introduction 96
4.2 Background 99
4.2.1 Text Mining and Natural Language Processing (NLP) 99
4.2.2 Text Mining Pipelines 101
4.2.3 Text Annotation 103
4.2.4 Corpus, Dictionary, and Pipeline Implementation 103
4.3 Description of Aim 2 105
4.3.1 Conceptual Model 105
4.3.2 Research Questions and Hypotheses 106
4.4 Methods 107
4.4.1 Corpora 107
4.4.2 Gold Standard Development 108
4.4.3 Pipeline Implementation 114
4.4.4 Evaluation Pipeline 115
4.5 Results 118
4.5.1 Gold Standard Creation 118
4.5.2 Dictionary and Pipeline Evaluation 119
4.5.3 Statistical Analysis 140
4.5.4 Dictionary and Pipeline 143
4.6 Discussion 146
4.6.1 Gold Standard Development 146
4.6.2 Error Analysis 148
4.7 Conclusion 154

5 PTSD DATA MINING 156
5.1 Introduction 156
5.1.1 Data Mining 156
5.1.2 Clustering of Clinical Data 157
5.1.3 Clinical Association Rules 158
5.2 Data Mining Background 158
5.2.1 Taxonomic Support 158
5.2.2 Cluster Analysis 159
5.2.3 Association Rule Mining 160
5.3 Description of Aim 3 161
5.3.1 Conceptual Model 161
5.3.2 Research Questions and Hypotheses 162
5.4 Methods 164
5.4.1 Dataset and Concept Identification 164
5.4.2 Data Mining Software 165
5.4.3 Cluster Algorithm Generation 165
5.4.4 Association Algorithm Generation 169
5.5 Results 170
5.5.1 Cluster Analysis 171
5.5.2 Association Rule Analysis 181
5.6 Discussion 188
5.7 Conclusion 191

6 DISCUSSION OF FINDINGS 193
6.1 Synthesis of Findings 193
6.2 Future Work 197

7 CONCLUSION 199
7.1 Summary and Lessons Learned 199
7.2 Closing Statement 200

REFERENCES 202
APPENDIX A Interview Questions 229
APPENDIX B Competency Questions 230
APPENDIX C Terminology Requirements 231
APPENDIX D Annotation Guidelines Subset 233
APPENDIX E Source Code – Gold Standard 234
APPENDIX F Source Code – Features 235
APPENDIX G Source Code – Core 237
APPENDIX H Source Code – Accuracy Metrics 242
APPENDIX I Association Rules 246
APPENDIX J Summary Accuracy Metrics 249

LIST OF FIGURES

Figure 1.1 Conceptual overview of dissertation project 6
Figure 2.1a General prevalence of PTSD 15
Figure 2.1b War-related prevalence of PTSD 15
Figure 2.2 Summary of types of terminology structures 18
Figure 3.1 Conceptual model of Aim 1 54
Figure 3.2 PTSD terminology needs assessment methodology 57
Figure 3.3 PTSD terminology requirements engineering methods 60
Figure 3.4 Agile requirement development 64
Figure 3.5 4-stage engineering development 66
Figure 3.6a PTSDO symptom parent classes 67
Figure 3.6b PTSDO treatment parent classes 67
Figure 3.7 Knowledge acquisition of PTSD terminology 71
Figure 3.8 cTAKES visual debugger of automated annotation 73
Figure 3.9 Example of concept or term reuse via IRI, xref, and CUI 74
Figure 3.10 Identification of end-user and stakeholder needs 75
Figure 3.11 Case report abstract 80
Figure 3.12 Post-processed case report abstract 80
Figure 3.13 PubMed concept discovery saturation curve 81
Figure 3.14 PTSD clinical guidelines concept discovery saturation curve 81
Figure 3.15 Selected symptom taxonomic relationships for PTSDO 85
Figure 3.16 Selected treatment classes 86
Figure 3.17 PTSDO term overlap 87
Figure 3.18 PTSDO concept overlap 88
Figure 4.1 Modular text mining pipeline for NLP tasks 99
Figure 4.2 Conceptual model of Aim 2 106
Figure 4.3 PTSD annotation schema and guideline development 111
Figure 4.4 F-score calculation 112
Figure 4.5a Recall formula 115
Figure 4.5b Precision formula 115
Figure 4.5c F-measure formula 115
Figure 4.6 Sample annotation from Semantator 118
Figure 4.7 PTSDO and PubMed symptom results 120
Figure 4.8 PILOTS and PubMed symptom results 121
Figure 4.9 SNOMED-CT and PubMed symptom results 122
Figure 4.10 NCI Thesaurus and PubMed symptom results 123
Figure 4.11 UMLS and PubMed symptom results 124
Figure 4.12 PTSDO and PubMed treatment results 125
Figure 4.13 PILOTS and PubMed treatment results 126
Figure 4.14 SNOMED-CT and PubMed treatment results 127
Figure 4.15 NCI Thesaurus and PubMed treatment results 128
Figure 4.16 UMLS and PubMed treatment results 130
Figure 4.17 PTSDO and psychotherapeutic transcripts symptom results 131
Figure 4.18 PILOTS and psychotherapeutic transcripts symptom results 132

Figure 4.19 SNOMED-CT and psychotherapeutic transcripts symptom results 133
Figure 4.20 NCI Thesaurus and psychotherapeutic transcripts symptom results 134
Figure 4.21 UMLS and psychotherapeutic transcripts symptom results 135
Figure 4.22 PTSDO and psychotherapeutic transcripts treatment results 136
Figure 4.23 PILOTS and psychotherapeutic transcripts treatment results 137
Figure 4.24 SNOMED-CT and psychotherapeutic transcripts treatment results 138
Figure 4.25 NCI Thesaurus and psychotherapeutic transcripts treatment results 139
Figure 4.26 UMLS and psychotherapeutic transcripts treatment results 140
Figure 4.27a Case report symptom metrics across dictionaries 144
Figure 4.27b Case report treatment metrics across dictionaries 144
Figure 4.28a Psychotherapeutic transcripts symptom metrics across dictionaries 145
Figure 4.28b Psychotherapeutic transcripts treatment metrics across dictionaries 145
Figure 5.1 Conceptual model of Aim 3 162
Figure 5.2 K-means concept clustering with UMLS 172
Figure 5.3a Scikit clustering supported with UMLS 173
Figure 5.3b DSM criterion of features 173
Figure 5.4 K-means DSM criterion clustering with UMLS 174
Figure 5.5 K-means concept clustering with PTSDO 176
Figure 5.6a K-Means clustering supported with PTSDO 177
Figure 5.6b DSM criterion of features 177
Figure 5.7 K-means DSM criterion clustering with PTSDO 178
Figure 5.8 EM concept clustering with PTSDO 179
Figure 5.9a EM clustering supported with PTSDO 180
Figure 5.9b DSM criterion of features 180
Figure 5.10 EM DSM criterion clustering with PTSDO 181
Figure 5.11 UMLS scatter plot of confidence, support and lift metrics 182
Figure 5.12 PTSDO scatter plot of confidence, support and lift metrics 183

LIST OF TABLES

Table 2.1 Groups utilizing PTSD terminology structures 16
Table 2.2 Applicable ontology development methods 27
Table 2.3 Clustering algorithms available for implementation 37
Table 2.4 Definitions for association rule mining 39
Table 3.1 Applicable agile methods for requirements development 50
Table 3.2 Attributes and meaning of SMART requirements methods 60
Table 3.3 Quality criteria for requirements analysis 62
Table 3.4 Development standards for PTSD terminology system 68
Table 3.5 Terminology resources for system population 70
Table 3.6 Subset of terminology requirements 77
Table 3.7 Collection of identified terms from various resources 82
Table 3.8a PTSDO symptom classes and synonyms 83
Table 3.8b PTSDO treatment classes and synonyms 83
Table 4.1 Low-level and high-level NLP tasks 100
Table 4.2 Summary of annotation guidelines 110
Table 4.3 Corpus for gold standard statistics 111
Table 4.4 Entity identification boundary definition 116
Table 4.5a Semantic types for symptom concept recognition 117
Table 4.5b Semantic types for treatment concept recognition 117
Table 4.6 Gold standard results 119
Table 4.7 PubMed case report symptom significant difference 142
Table 4.8 PubMed case report treatment significant difference 142
Table 4.9 Psychotherapeutic transcripts symptom significant difference 145
Table 4.10 Psychotherapeutic transcripts treatment significant difference 146
Table 4.11 False-negative PubMed symptom error analysis 149
Table 4.12 False-positive PubMed symptom error analysis 150
Table 4.13 False-negative PubMed treatment error analysis 150
Table 4.14 False-positive PubMed treatment error analysis 151
Table 4.15 False-negative psychotherapeutic transcripts symptom error analysis 151
Table 4.16 False- positive psychotherapeutic transcripts symptom error analysis 152
Table 4.17 False-negative psychotherapeutic transcripts treatment error analysis 152
Table 4.18 False- positive psychotherapeutic transcripts treatment error analysis 153
Table 5.1 Transaction extracted concepts via SKATE supported with PTSDO 166
Table 5.2 UMLS concepts corresponding to trauma exposure 182
Table 5.3 PTSDO concepts corresponding to trauma exposure 183
Table 5.4 Association rules obtained from PTSDO 185

ACKNOWLEDGEMENTS

I would like to thank my advisor and committee members:

- Advisor: Aaron Cohen, MD
- Committee Chair: Melissa Haendel, PhD
- Thomas Rindflesch, PhD
- Steven Bedrick, PhD
- Jianji Yang, PhD
- Blake Lesselroth, MD

I would also like to acknowledge the following groups and institutions that has provided invaluable support:

- Department of Medical Informatics and Clinical Epidemiology (DMICE), Oregon Health & Science University (OHSU)
- National Library of Medicine (NLM)
- OHSU Ontology Development Group
- Semantic Knowledge Representation Group, NLM
- Department of Veteran Affairs Healthcare System
- Yale School of Medicine, Yale University

I appreciate the support from my family including my wife and son. I also thank the many friends at OHSU and DCIME whom I have had collaboration with, who have supported me, contributed to the completion of this dissertation, and made my research that much more enjoyable.

This research was supported by the National Library of Medicine of the National Institutes of Health under Award Number T15LM007088 and by the National Science Foundation under grant number 1111722. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or National Science Foundation.

ABSTRACT

OBJECTIVE: Information overload in the biomedical domain is a well-researched problem of the last several decades but its pace of expansion has increased exponentially. Continued development of tools and applications will be required to combat the growing collection of data. To obtain maximum accuracy, text mining tools supported with robust biomedical standardized terminologies are necessary implementations congruent with these application development initiatives. This dissertation focuses on developing a terminology system specifically designed to aid in the extraction of post-traumatic stress disorder (PTSD) concepts from unstructured resources. This research draws upon several natural language processing (NLP) techniques, specifically improving named entity recognition (NER) in a multi-stage pipeline approach to increase coverage of the PTSD medical sub-domain. Based upon a rigorous requirements gathering process, the design and development of a PTSD terminology system (PTSDO) is described for implementation of entity identification tasks. In order to evaluate the vocabulary coverage and pipeline accuracy, two annotated corpora to serve as concept gold standards are created. Once salient concepts are extracted, data mining algorithms of cluster and association rule analyses are implemented to discover potentially useful and previously unknown information from collections of data. METHODS: Methodologies from the fields of software development and ontology engineering, supplemented with agile methodology techniques, are implemented for terminology development. The building process consists of iterative steps, namely scoping, reusing existing terminology resources, enumerating required terms, then defining the hierarchical arrangement. After knowledge is acquired from all available resources, semi-automated concept and term mining iterative methods are applied to unstructured domain-related corpora to identify salient entities for input into PTSDO. To evaluate coverage, gold standards are developed following standards established for NLP tasks in order to maximize results oriented annotation. Four text mining pipelines are implemented and supported with five terminology resources to evaluate coverage and significant difference. Based upon extracted entities supported with the UMLS, those extracted with PTSDO support are compared in the implementation of data mining algorithms. Clustering analysis among symptom concept occurrence probabilities are calculated from a co-occurrence matrix to discover insights into the

data points. Association rule mining using the apriori algorithm developed from the extracted concepts provide useful insights into correspondence between PTSD symptomatology. RESULTS: Needs assessment and requirements gathering determined a prerequisite for identification of named entities, adherence to system design standardization, and text mining support specification. The knowledge acquisition techniques identified 3516 symptom candidate terms and 2008 candidate treatment terms for input into PTSDO. Concept input into the terminology framework includes a total of 509. Term input includes a total of 2582 consisting of alternative strings of text, synonyms, acronyms, and abbreviations. Concept overlap ranges from 72.5% to 84.85% and term overlap ranges from 58.83% to 65.23%. The gold standard development resulted in high inter-annotator agreements (overall F-measures between 0.80 and 0.91) for the annotation of case reports and between (0.68 and 0.84) for psychotherapeutic transcripts. The accuracy of the text mining pipelines supported with the terminologies exhibited significantly different metrics and their respective error analysis provided insight into strengths and limitations. PTSDO obtains a maximum F-measures of 0.94 (0.95 precision and 0.92 recall) for case report concept recognition and a maximum F-measures of 0.91 (0.86 precision and 0.95 recall) for psychotherapeutic transcripts. Clustering and association rule learning using extracted concepts from a text mining pipeline are compared to a baseline supported with the UMLS to extracted concepts supported with the PTSDO. Findings from the cluster analysis show that concepts group more consistently between clusters when supported with PTSDO in comparison to the UMLS. The association rules developed with the extracted concepts provide useful insights into correspondence between PTSD symptomatology. Analysis of association rule mining supports hypothesis generation via the identification of relationships between symptoms that merit further investigation. CONCLUSION: This research provides a synthesized terminology for clinicians, researchers, and developers who need to share information about PTSD. Utilizing a domain-specific terminology framework to support text mining pipelines produce highly accurate annotations beneficial to training more advanced algorithms. Text mining and concept recognition systems require significant efforts to improve usability. Highly informational data mining algorithms can identify useful patterns in data with the support of the accuracy improvements from pipeline and terminology development. By increasing the semantic processing capability of PTSDO with increased properties and formal logic, its enhancements can provide support for advanced clinical decision support applications of the future.

Preface

This chapter provides an overview of the problem space this dissertation addresses in Section 1.1, and the significance of the research in Section 1.2. Section 1.3 lists the primary research questions followed by an overview of each of the project aims. Section 1.4.1 describes aim 1 and the development of a PTSD terminology system to include the research hypotheses. Section 1.4.2 introduces the implementation of several text mining pipelines for aim 2 to extract PTSD information supported with the newly developed terminology system compared to robust existing resources. Section 1.4.3 introduces aim 3, the implementation of several data mining algorithms driven by concepts extracted in aim 2 and supported with the terminology developed in aim 1. Section 1.5 discusses the main contributions of this research and Section 1.6 outlines the structure of the dissertation.

Chapter 1

Introduction

1.1 Aggregated Terminology for PTSD

The importance of standardizing healthcare terminology has long been established to provide data structure for synthesis and sharing across the continuum of care [1, 2]. Utilizing terminology frameworks such as controlled vocabularies, taxonomies, and ontologies promote standards and improve communication by providing a formal representation of the entities in a specified domain [3, 4]. However, the utility of these terminologies in clinical or health information applications has not yet been fully demonstrated. In this paper, we describe the development of a terminology framework in the medical sub-domain of post-traumatic stress disorder (PTSD). The American Psychiatric Association defines PTSD as a condition occurring from exposure to a trauma that impacts the physical integrity or life of the individual or of another person [5, 6]. It is considered normal for an individual to have a strong reaction to a traumatic event but the effects should decrease over time when the threat is no longer present. However, people with PTSD continue to experience extreme reactions and symptoms even after the trauma is no longer present [7]. According to the National Center for PTSD, 7-8% of the population in the U.S. will have a form of this disorder at some point in their lives [8, 9].

The prevalence of PTSD continues to grow, particularly in the military veteran population, as combat operations continue around the globe and as researchers begin to better understand the disorder and its presence in patients once previously missed [7, 10]. The current healthcare system is also not equipped to cope with this disorder. The continuous stream of PTSD evidence is outpacing the ability to process such data. Specifically, the challenge of the data stored in narrative free text contributes to the complexity of dealing with the disorder. From an informatics standpoint, there are significant opportunities to assist healthcare communities with improved PTSD understanding through initiatives of health information system development, information extraction (IE) and natural language processing (NLP) tools and projects. Before these tools and applications are built, standards and interoperability and how to maximize

collaboration must be examined. Consideration must be given to the various ways PTSD terms are defined and used in healthcare and the research community.

The availability of knowledge bases has been identified as one of the critical elements that can help to realize the benefits of information management and clinical decision support [11]. This is especially true for PTSD vocabulary that is subjective, ambiguous, and overlapping with other mental health disorders [7]. There is a lack of collective terminology necessary to support the future of clinical applications, and informatics initiatives in the domain. Improvements to vocabulary must be made to address terminology issues such as data heterogeneity (various data formats and structures), disambiguation (multiple word meanings), missing and incomplete information. These improvements to develop a PTSD terminology framework requires a rigorous needs assessment to identify vocabulary and coverage requirements. The goal is to discover important user needs that can be translated into knowledge acquisition requirements. The requirements engineering process consists of specifications that define, describe, and unambiguously communicate the stakeholder needs [12]. This dissertation explores terminology development under the assumption that inadequate requirements engineering would inhibit quality of the overall project. The terminology development methods described in this paper are based on various stages of complexity involved within vocabulary synthesis. The aim of the methods is to provide the non-expert developer with a framework capable of supporting terminology application in multiple use cases and at varying levels of complexity. Appropriate documentation fosters transparency, traceability, and replication of the development steps [13].

High quality terminologies facilitate data aggregation and can expedite translating research into implementation [14]. Text mining methods on data support processing into knowledge functional in clinical practice and medical science research. These methods are most efficient when supported with terminology systems built with standards and sufficient coverage for explicit use cases [14, 15]. As stated by Cohen and Hersh, modest text-matching algorithms are insufficient for named entity recognition without a complete domain dictionary [16]. The foundation of even the most sophisticated text mining systems depend upon the identification of domain specific concepts and terms with accurate entity identification. Considering the degree that coverage is maximized when supported by a comprehensive dictionary source, the development of this framework is priority due to existing inadequate resources.

1.2 Problem Statement

While the prevalence of PTSD continues to grow, it is still largely under-detected due to the difficulties in diagnosis, volatility of symptoms and lack of effective screening in healthcare facilities. The insufficiency of adequate screening is an important public health initiative to be addressed [7, 10]. The current U.S. healthcare system is not equipped to adequately cope with this disorder.

A major gap inhibiting the research-to-cure time frame for PTSD is that many of these initiatives are operating in silos instead of cross functional collaborations. These initiatives have many of the same goals, share the need for better access to high quality information, and must overcome interoperability issues. A large percentage of data used by these initiatives and similar research is stored in narrative free text that is highly unstructured [17]. From an informatics standpoint, there are significant opportunities to assist healthcare communities with improved PTSD understanding through initiatives of health information system development, and through information extraction (IE) and natural language processing (NLP) tools and projects [18, 19].

Without proper use of terminology, it can become a bottleneck for the deployment of innovation and the assessment of quality in PTSD health care [20]. Opening that bottleneck will require understanding of textual elements and leveraging it into an asset for health care and research. A significant challenge to leveraging the information that can be developed from this data are the many different formats. Most existing and independent databases have overlapped data as they are built remotely and separate. All of this information captured in different formats can challenge the understanding of existing relationships between different data. Many researchers need much of the same data but with different meaning or context [21].

Clinical and research data are distributed across various heterogeneous databases structures making the sharing of information difficult if not impossible [22]. Individual database development creates dissimilar data formats, making integration costly in both monetary and time-based terms. Such barriers add to the challenge of retrieving information using common language in an understandable form [23]. Its manual accumulation is time-consuming and propagates tedious tasks of searching that are not feasible for working clinicians and researchers. The efficacy of readily accessible text mining systems is hampered by unavailable terminology systems and insufficient annotated training data [24]. This dissertation addresses the task of

using text mining for the aim of concept discovery and implements a data mining use case in order to show the applicability of terminology development for improved accuracy.

1.3 Significance and Motivation

An aspect at the heart of medical informatics is the role of an informatician as being the bridge between information technology, medicine, library services, and the endless lists of sub-fields that need unification. Towards the development of clinical decision support systems, healthcare's ultimate reliance on automated suggestions, user acceptance will be determined by their respective accuracy. It is exciting but difficult to predict the biomedical applications informaticists will build in the future, but they will need to be supported by various types of knowledge bases and terminology structures in order to achieve the accuracy necessary. These terminology backbones have the opportunity to be the bridge between unstructured data and structured data. The motivation is to build a terminology representation in a standard format accessible for computer algorithms to interpret the data it contains. With the compounding expansion of information overload, automation techniques are key to overcoming its unhampered growth. A terminology system can directly support biomedical applications or be utilized for automated annotation to build training data sets. This training data is necessary for other text mining strategies to generalize or make various predictions about new data [25]. Garnering relevant information from a plethora of medical and life science resources provides benefits to multiple stakeholders. The ability to extract and provide meaning to unstructured data yields insights that may otherwise not be obtainable or obvious. It is the goal for this terminology system to support medical text mining as a prelude to data mining and clinical decision support.

1.4 Primary Research Questions

The focus in this dissertation is on the development and use of a terminology framework to support dictionary and rule-based entity identification of PTSD concepts. This tool is capable of extracting entities from resources such as clinical progress notes, published literature, and online health information to populate a PTSD vocabulary framework. The problem statement discussed above in the motivation section inspires the tasks for this research and the goals this dissertation set about to accomplish. This dissertation focuses on terminology and vocabulary development

to support identification of concepts necessary to accurately depict language used in the description of this disorder. The identification process is iterative in that it implements a semi-automatic approach where, as new terms are added to the domain from novel research or available phenotype information and missed by the system are added to the vocabulary. The research questions addressed in this dissertation span the following objectives toward the development and implementation of a PTSD terminology system: 1) needs assessment, vocabulary knowledge acquisition, requirements and terminology engineering; 2) text mining pipeline (natural language processing) implementation for examining concept coverage and information extraction, and 3) clustering and affinity analysis of signs and symptomatology. Fundamental to meeting the goals of the dissertation, the following two primary research question will be addressed in subsequent sections:

1. *Can sufficient PTSD knowledge be acquired in order to build a terminology framework capable of sharing information among stakeholders and utilized to support a diverse set of biomedical applications?*

2. *Can PTSD concepts be accurately extracted utilizing a hybrid dictionary and rule-based approach to support concept recognition and data mining applications?*

Figure 1.1 shows the interconnection of the two primary research questions, and a broad overview of their implementation in real world scenario-based biomedical applications. This dissertation addresses variations of concept mining objectives with three primary goals: *(a) identifying terms, concepts, and synonyms in text, (b) annotating meta-data to include definitions, source vocabulary, and examples of usage,* and *(c) assigning DSM (Diagnostic and Statistical Manual of Mental Disorders) categories to these concepts*. Furthermore, the framework includes methods for automatically verifying and linking entities to published ontologies, knowledge bases, and terminology structures. The two primary research questions are elaborated upon and further divided into five subsequent questions described below. These more specific research questions are attached to their associated aims and detailed in order to meet project objectives.

1.5 Overview of Research Aims

The conceptual overview of this dissertation is shown in **Figure 1.1** which begins with a collection of PTSD terminology that requires knowledge acquisition from salient resources. The accumulated concepts and terms are stored in an extensible markup language format that can transformed into various dictionary-based formats in order to support a diverse availability of text mining platforms. The newly developed terminology system coverage is validated by its support of these text mining natural language processing (NLP) tools and compared to existing available lexical resources. While both rule-based and statistical learning methods unique to each NLP tool enhance results, the equalization of the differing dictionaries make the analyses comparable. The use of the developed terminology focuses on a dictionary-based approach to improve the NLP task of named entity recognition for concept identification that can iteratively be processed for adding missing terminology for system population. Via the accumulation of extracted concepts, this collection can be applied to a multitude of biomedical applications for which this research focuses on tasks in data mining.

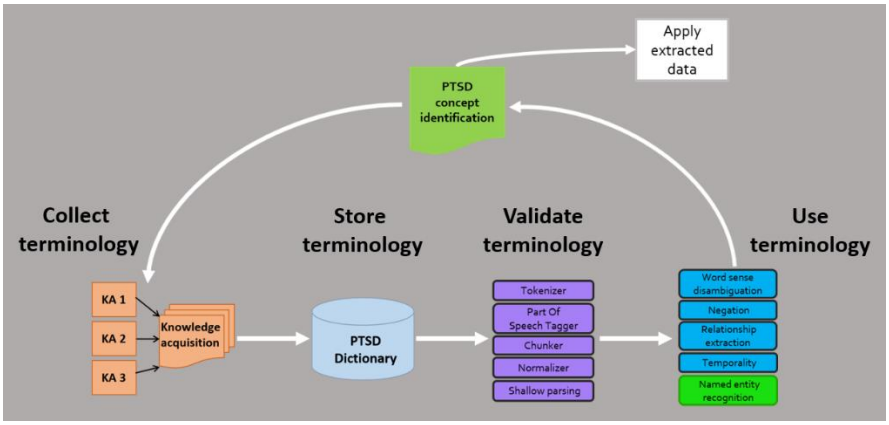


Figure 1.1. Conceptual overview of dissertation project

1.5.1 Aim 1

Assess PTSD terminology needs for the requirements, knowledge acquisition, and development of a terminology framework to support text mining tasks

1.5.1.1 Aim 1 Research Questions

The coverage and feasibility of implementing existing resources in order to achieve stated projects goals is examined. The specific collaborative knowledge acquisition approaches

discussed in subsequent chapters is detailed for facilitating knowledge transfer among current and future research projects. Verification of extracted information is one of the key processes to address the quality of concepts in the terminology framework in addition to addressing the reusability of published data. Entities, concepts, and terms detailed with associated metadata and complementary extracted information supports removing ambiguity within the terminology framework.

RQ1. *Can sufficient PTSD terminology resources be identified in order to acquire salient concepts for expansion of domain coverage?*

Task 1: Perform manual mining of existing terminology resources

Task 2: Assess needs of stakeholders through iterative interviewing for requirements vetting and knowledge acquisition from experts

Task 3: Implement text mining pipeline for semi-automated mining of terms and concepts

Task 4: Examine knowledge acquisition saturation of concepts, terms and words from a range of terminology resources

RQ2. *Can the identification of PTSD entities described implicitly in unstructured data be automated in an iterative method to populate a structured hierarchical framework?*

Task 1: Examine concept overlap of acquired knowledge from available resources

Task 2: Examine term overlap of acquired knowledge from available resources

Task 3: Perform iterative domain expert validation of collected concepts and terms

1.5.1.2 Aim 1 Research Hypotheses

Hypothesis 1 - Manual mining of existing terminology resources performed during needs assessment will produce a generic PTSD terminology system base to maximize stakeholder input.

Hypothesis 2 - Manual knowledge acquisition of existing terminology resources will not produce the comprehensive needed concept coverage.

Hypothesis 3 – An iterative semi-automated text mining approach to knowledge acquisition will determine when sufficient coverage is met in order to satisfy identified stakeholder needs.

Hypothesis 4 – Concept overlap will exist by more than 50% in existing concepts identified from available resources.

Hypothesis 5 - Term overlap will exist by more than 50% in existing concepts identified from available resources.

1.5.2 Aim 2

Implement text mining pipelines supported with biomedical terminology resources for validation of PTSD concept coverage

1.5.2.1 Aim 2 Research Questions

With the growing use and development of terminology resources, evaluation is important in helping to determine their respective suitability for the requirements of an application or domain of usage. A combination of terminology resources and dictionary-based tools are applied to test the hypotheses associated with this aim.

RQ3. *Do terminology resources containing PTSD concepts vary among text mining pipelines when processing biomedical corpora?*

Study 1: Automated annotation accuracy evaluation of symptom concepts in PubMed PTSD case reports with each dictionary and pipeline combination.

Study 2: Automated annotation accuracy evaluation of treatment concepts in PubMed PTSD case reports with each dictionary and pipeline combination.

Study 3: Automated annotation accuracy evaluation of symptom concepts in psychotherapeutic therapeutic session transcripts of PTSD patients with each dictionary and pipeline combination.

Study 4: Automated annotation accuracy evaluation of treatment concepts in psychotherapeutic therapeutic session transcripts of PTSD patients with each dictionary and pipeline combination.

1.5.2.2 Aim 2 Research Hypotheses

Hypothesis 1 - Dictionary-based terminology resources will not perform equally on corpora with text processing pipelines.

Hypothesis 2 – The accuracy of dictionary coverage will significantly vary between text processing pipelines.

Hypothesis 3 – The accuracy of PTSDO will be significantly higher than coverage of other evaluated dictionaries with text processing pipelines.

1.5.3 Aim 3

Utilize extracted PTSD signs and symptomatology concepts to perform clustering and affinity analysis for pattern recognition mining

1.5.3.1 Aim 3 Research Questions

These research questions explore whether the co-occurrences of PTSD concepts can be exploited in order to gain insight into the data points. The UMLS (Unified Medical Language System) is implemented as a baseline for comparison of results with the developed PTSD terminology system from aim 1.

RQ4. *Do PTSD symptom concepts group consistently between clusters in accordance with their hierarchical representations from DSM diagnostic criterion?*

Study 1: A cluster analysis is conducted in order to categorize the co-occurrence of concepts that group together and identify the concepts within each cluster with the highest probability of occurrence.

Study 2: Each concept's groupings are analyzed according to their corresponding DSM diagnostic criterion to explore its respective dispersion for each cluster.

Study 3: The formed clusters to include the concept co-occurrences supported with the PTSD terminology system is compared to the clusters and concept co-occurrences extracted using the UMLS as a base comparison.

Study 4: The data points in each cluster is categorized according to its DSM diagnostic criterion to include the concept co-occurrences and the uniformity of the arrangements supported with PTSDO is compared to those supported with the UMLS. The clusters developed from each terminology source is expected to disperse concepts from each of the DSM diagnostic criteria.

RQ5. *Can association rule mining produce insights into PTSD symptomatology to support hypothesis generation?*

Study 1: Association rule mining is conducted to in order to identify concepts with high degrees of association over a set threshold.

Study 2: The metrics of exact associations rules produced by extraction supported with the PTSDO and the UMLS are compared.

Study 3: Adjustments to minimum threshold is adjusted in order to examine effects on the type and strength of corresponding associations obtained.

1.5.3.2 Aim 3 Research Hypotheses

Hypothesis 1 – The most commonly identified symptoms among PTSD patients will make up the co-occurrences of formed clusters with the highest probabilities

Hypothesis 2 – There will be a difference in the dispersion of concepts in the clustering supported with the PTSDO compared to the UMLS

Hypothesis 3 – The clusters formed will produce concepts that group uniformly with the DSM diagnostic criterion

Hypothesis 4 – The concepts extracted using the UMLS will produce a larger number of threshold met associations than the concepts extracted using the PTSDO

Hypothesis 5 - There will be no difference in the association metrics produced with the PTSDO and those from the UMLS

1.6 Contribution

This dissertation describes the synthetization of terminology resources to improve text processing pipelines for PTSD entity identification in order to populate symptom characteristics and treatment categories with new and existing concepts for data mining applications. A contribution of this dissertation is the provision of an explicit example of terminology development in this highly complex medical sub-domain in mental health. The problem of terminology expansion and integration from distributed databases and existing PTSD resources fosters a better understanding of the complexities involved. An approach is demonstrated for acquiring knowledge in order to populate a knowledge base for this highly specialized domain. A semi-automated technique for extracting entities and entity information from unstructured

sources is presented to provide structure by categorizing recognized concepts. This approach highlights the value of an approach focusing on a dictionary-based approach to improve entity recognition. To address the issues of ambiguity and missing information about entities, this dissertation provides an approach to exploits natural language documents as an additional source of metadata. A simple experiment to discuss the usefulness of the DSM categorization and association rule mining for hypothesis generation is detailed to drive home the value of text mining supported with the efforts of terminology development.

1.7 Dissertation Structure

The remainder of the dissertation is divided into seven chapters and structured as follows. Chapter 2 provides the background and significance of terminology resources, natural language processing (NLP) frameworks, and data mining applications applicable to the biomedical literature and clinical data. Chapter 3 introduces aim 1 and a thorough needs assessment for requirements engineering in the development of a PTSD terminology system. The results of rigorous requirements gathering and the knowledge acquisition of concepts and terms from multiple lexicons, domain expert interviews, and a semi-automated text mining method. Chapter 4 describes aim 2 and the evaluation of implementing multiple text mining pipelines supported with various dictionary resources. Details are presented regarding the extraction of concepts accuracy metrics and significant difference comparisons between the pipelines and terminologies as dictionaries combinations. Chapter 5 presents aim 3, a data mining implementation from the most accurate results captured from aim 2. A cluster analysis is performed based upon the probability of concept co-occurrence to which each concept is translated to its DSM (Diagnostic and Statistical Manual of Mental Disorders) criterion for further analysis. The chapter closes with the analysis of an association rule mining study for the discovery of interesting patterns in the concept co-occurrences. Chapter 6 discusses the findings across the three aims including the results of the text mining pipelines and error analysis. The clustering results, association rules and their potential impact and limitations of the project aims are conferred. Chapter 7 summarizes what has been learned from this research and future investigations to expand the relevant findings. The chapter closes with potential directions for future work such as the applications of developing terminology resources for text mining initiatives in other areas of mental health and medicine.

Chapter 2

Background

2.1 Outline

This chapter provides the background for the aims presented in the introduction. It is important to recognize aspects of post-traumatic stress disorder (PTSD) that make the disorder difficult to understand, as well as the tools for aggregating and synthesizing a collection of PTSD terminology. The structural formats available to represent the disorder, and the accessible biomedical resources from which to collect concepts and terms, specifically those within the mental health domain, will be reviewed. Lastly, there are fundamental frameworks and structures that have been used to build text mining applications and complete data processing tasks, which will be outlined in this chapter. This section closes with a summary of gaps in knowledge of PTSD terminology structures, medical and mental health terminology resources, text mining, and data mining opportunities.

2.2 Post-traumatic Stress Disorder (PTSD)

The American Psychiatric Association defines post-traumatic stress disorder (PTSD) as a condition occurring from exposure to a trauma that impacts the physical integrity or life of the individual or of another person [5, 6]. It is considered normal for an individual to have a strong reaction to a traumatic event but the effects should decrease over time when the threat is no longer present. However, people with PTSD continue to experience extreme reactions and symptoms even after the trauma is no longer present [7].

Although PTSD is a relatively new psychiatric disorder, first occurring in 1980 in the DSM-III, the construct of the disorder has had a long history referred to by other names. Currently, the disorder is described by the psychiatric diagnosis of PTSD. Previously, it was described by the event that caused it such as shell shock, railway spine, war neurosis, concentration-camp syndrome and rape trauma syndrome [26, 27]. Each of these experiences resulted in symptoms that are very similar to the current variations of PTSD symptomatology today [28]. These

symptoms generally include: 1) reliving or re-experiencing the traumatic event; 2) avoiding situations that trigger memories of the event; 3) negative changes in beliefs and feelings; and 4) feelings of hyperarousal [29-30].

Treatments for PTSD consist of many psychological and social interventions such as stress management and behavior based therapy [31-32]. Psychotherapy helps the patient recover a sense of self, learn new coping strategies and ways to deal with intense emotions related to the trauma [30, 33]. Medications prescribed are guided by pronounced variations in symptomatology. The selective serotonin reuptake inhibitors (SSRIs) appear to help the core symptoms when given in higher doses for five to eight weeks, while the tricyclic antidepressants (TCAs) or the monoamine oxidase inhibitors (MAOIs) are most useful in treating anxiety and depression [30, 34]. Some alternative therapies for PTSD include various forms of meditation, yoga, religious counseling, art and dance therapy [35-36].

2.2.1 PTSD Prevalence

As shown in **Figure 2.1a**, 70% of adults in the United States have experienced at least one traumatic event in their lives [8, 37]. Additionally, 7-8% of adults have been diagnosed with PTSD at least once in their lives [38]. Between 60-80% of the people who experience a severe traumatic event get post-traumatic stress disorder [9]. More than 50% of the PTSD victims in the United States become alcohol dependent [7]. Prevalence for war-related disorder is higher and findings going forward from Vietnam is shown in **Figure 2.1b**. There is data available from World War I and II, however its validity is questioned due to collection methods and outdated tests. Studies vary but indicate that anywhere from 10-30% of all combat veterans will suffer from the disorder at least once in their lives. It is estimated that around 84% of the Vietnam veterans diagnosed with PTSD suffer from at least moderate impairment today [39-40]. Roughly 20% of the military personnel returning from Afghanistan or Iraq have the disorder [10, 41]. War veterans with PTSD more often get divorces, become single parent and/or homeless. The effects of the disorder can lead to other serious health problems such as alcohol abuse, depression, or suicide [8, 9].

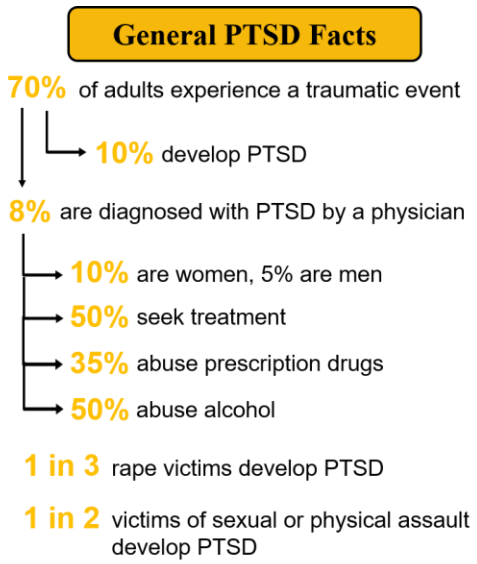


Figure 2.1a. General prevalence of PTSD

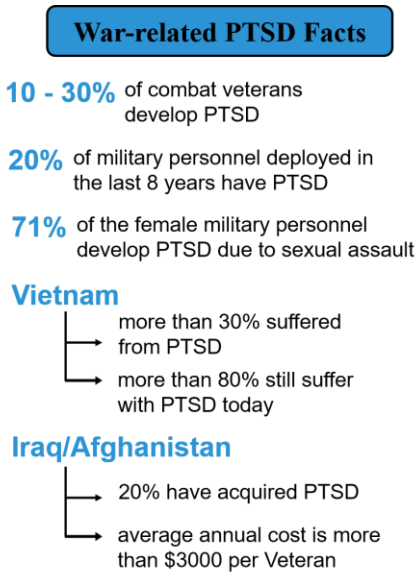


Figure 2.1b. War-related prevalence of PTSD

2.2.2 Initiatives involving PTSD Terminology

There are currently many ongoing projects, some of which are shown in **Table 2.1**, that are either accumulating, using, establishing, or in need of standard terminology for PTSD. Various organizations and initiatives have an interest in using a robust knowledge base for interoperability with other mental health projects around the globe. The Veteran Affairs (VA) healthcare administration has been a leader in PTSD projects and initiatives ongoing as well as providing information to both patients and professionals at www.ptsd.va.gov. The Department of Defense (DoD) support initiatives within the Defense Centers of Excellence for Psychological Health and Traumatic Brain Injury (DCoE), comprised of Defense and Veterans Brain Injury Center (DVBIC) [42], Deployment Health Clinical Center (DHCC) and National Center for Telehealth and Technology (T2). T2 serves as the principal integrator and authority on psychological health and traumatic brain injury (TBI) knowledge and standards for the DoD. DVBIC has created and maintains a TBI surveillance database and has been executing a 15-year longitudinal study of the effects of TBI in Operations Iraqi and Enduring Freedom service members and their families [42-43]. DHCC develops evidence-based treatments and clinical support tools and integrates behavioral health data for identification and treatment of psychological health activities [44-45]. The Military Suicide Research Consortium (MSRC) is part of an ongoing strategy to integrate and synchronize U.S. Department of Defense and civilian

efforts to implement research on suicide prevention [46]. U.S. Department of Health and Human Services operates a vast network of sharing information as well as the International Society for Traumatic Stress (ISTSS). ISTSS is dedicated to the discovery and dissemination of knowledge about policy, program and service initiatives that seek to reduce traumatic stressors and their immediate and long-term consequences [47-49].

- American Psychiatric Association
- Anxiety Disorders Association of America, Inc.
- International Critical Incident Stress Foundation, Inc.
- International Society for Traumatic Stress Studies
- National Center for PTSD
- National Institute of Mental Health
- U.S. Department of Health and Human Services
- VA Consortium for Healthcare Informatics Research (CHIR)
- Common Data Element (CDE) Project
- One Mind GEMINI Program
- Defense Centers of Excellence for Psychological Health (DCoE)
- Department of Defense Task Force on Mental Health
- The Defense and Veterans Brain Injury Center (DVBIC)
- The Military Suicide Research Consortium (MSRC)

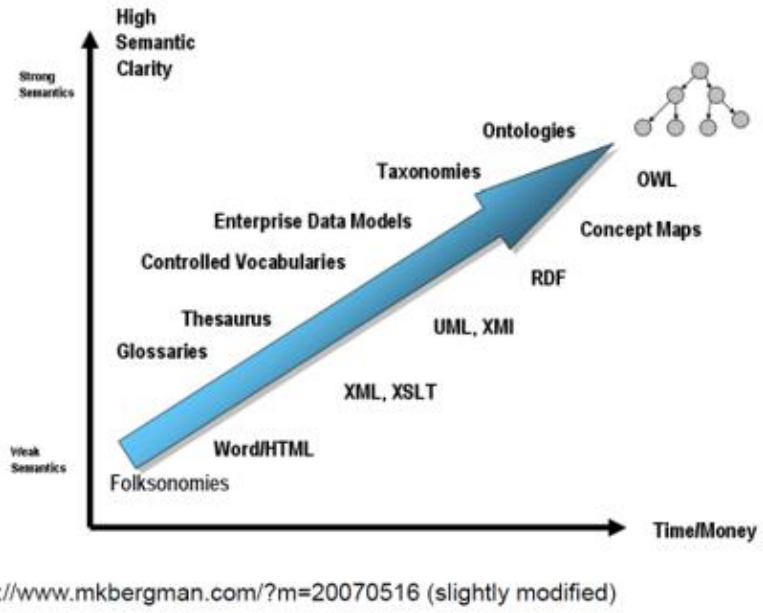
Table 2.1. Groups utilizing or in need of PTSD terminology structure [42-57]

There are many ongoing projects that have a direct effect on terminology in PTSD. The VA's Consortium for Healthcare Informatics Research **(**CHIR) [50] is a multi-disciplinary group of collaborating investigators formed to advance the use of unstructured text and other types of clinical data in the EHR and supports projects to understand how clinicians document clinical information about PTSD patients [51]. Because a multitude of data formats creates barriers to data sharing, the Common Data Element Project [52] is an initiative to identify common data elements (CDEs) used in clinical research [53]. As a part of this project, a workgroup has been developing CDEs for PTSD for improved analysis and sharing across the research community [54]. One Mind is a non-profit organization whose mission is to take the lead role in the research, funding, marketing, and public awareness of mental illness and brain injury. This organization wants to promote policy improvement that accelerates the research-to-cure time frame for TBI and PTSD and are currently exploring knowledge representation possibilities for the data they are accumulating [55]. Clinical Data Interchange Standards Consortium (CDISC) [56] is a global, open, multidisciplinary, non-profit organization that has established standards to support the acquisition, exchange, submission and archive of clinical research data and metadata. The CDISC mission is to develop and support global, platform-independent data standards that

enable information system interoperability to improve medical research and related areas of healthcare [57].

2.3 Terminology Structures and Formats

As stated by Spasic et al., term variation initiates from "the ability of a natural language to express a single concept in a number of ways" and a high number of synonyms [58]. For example, in biomedicine there are many synonyms for proteins, enzymes, genes, etc. [59]. Acronyms are also used at length recognized by Chang et al. which stated that "a new acronym is introduced in every five to ten abstracts" [60]. Additionally, ambiguity is inherent in clinical medicine and life science text. Spasic et al. stated this is because the "evolution of species fosters homologues and analogues" [58]. Terminology formats developed via aggregation and synthesis provide structure to the uncertainty inherent in natural language. A terminology consists of a collection of terms, words, and phrases representing and describing a system of concepts within a specific domain [61]. Various formats such as those displayed in **Figure 2.2** have been developed to support a range of applications. As shown along the spectrum developed by Lee Obrst [62], the cost of time and money increase proportionally with the semantic processing ability of each complex structure. A general rule in order to apply the most beneficial cost/benefit ratio is to only build a terminology structure with enough semantic processing ability needed in order to support its intended application. Terminologies and coding systems in medicine date back many years. The International Classification of Diseases is one of the first standard terminologies whose lineage traces back at least to 1893 [61, 63]. The complexity and thoroughness of clinical terminologies has increased since biological and scientific classification schemes were first developed [21]. The framework of a robust design must adequately be represented with coverage as described by its specified development scope. There are numerous published clinical terminologies with general medical domain and explicit sub-domain coverage. However, the quality of usability varies with implementation needs and purpose of conceptual design.



http://www.mkbergman.com/?m=20070516 (slightly modified)

Figure 2.2. Summary of types of terminology structures [62]

A thesaurus is a networked collection of controlled vocabulary terms. This means that a thesaurus uses associative relationships in addition to parent-child relationships. The expressiveness of the associative relationships in a thesaurus vary but typically has two kinds of links: broader and narrower term [64-65]. Webster's defines a controlled vocabulary as a "predefined linguistic representational list of structured terms used to describe domain-specific concepts" enumerated explicitly [65]. All terms in a controlled vocabulary should have an unambiguous, non-redundant definition. It typically includes preferred and alternative terms and provides a means of conveying scientific and technical information [66] including definitions agreed by subject matter experts or a research community [67]. Controlled vocabularies are important in this research because there are many ways to say the same thing in the natural language of PTSD terminology [68]. For example, abnormal sleep pattern, hypersomnolence, insomnia, and sleeplessness all refer to the concept dyssomnia. However, hypersomnolence is defined as "excessive sleep" where insomnia is "a condition of not being able to sleep" [8-9]. A controlled vocabulary removes the ambiguity inherent in natural language. It supports matching user language with that used to formally describe, organize, and categorize data in information resources such as journal articles and books [69].

Cimino's desiderata [70] described the design of a controlled healthcare vocabulary in order to carefully consider the variations in purpose when modeling the terminology framework. This

work recognized the challenges in developing a structure for multi-purposes and vocabulary considerations individual to the purpose of one's design. Relevant desiderata applicable to this research are considerations focused on content coverage, concept meaning, classification, definitions, level of granularity, focus on changing requirements, and a means to recognize redundancy. Cimino's list is intended to serve as a discussion starter for developing and structuring terminology representation [70].

A taxonomy is a controlled vocabulary in which all terms are classified into ordered categories with a hierarchical structure. The structure will have a parent/child or broader/narrower relationship or both to other terms in the taxonomy [71]. Hierarchical representation serves as the basis from which vocabulary developers build robust and reliable terminologies. Current clinical terminologies develop hierarchies of information to create specializations of information. The major hierarchies within most terminologies consist of the "is-a" and "part-of" relations which serve as the basis for concepts to inherit properties from multiple hierarchies. A taxonomy structure is often characterized as a 'tree' where its terms are referred to as 'nodes.' A node may be repeated at more than one place within the taxonomy if it has multiple broader terms. Some taxonomies allow poly-hierarchy, which means that a term can have multiple parents. This means that even if a term appears in multiple places in a taxonomy, it is still the same term. Specifically, if a term has children in one place in a taxonomy, then it has the same children in every other place where it appears [21].

Requirements of granularity led to medical classification systems which are organized structures for arranging or classifying concepts based on similar characteristics [72]. A classification system aggregates data at a prescribed level of abstraction for a particular domain fostering improved consistency of conceptual data [73]. The classification and categorization of concepts have advanced with assertional knowledge and logical definitions as complexity of terminology structures has increased [74]. These structures utilize formal definitions to categorize concepts and predicates its related definitional information [21].

In the domain of computer science, Gruber defines an ontology as an "explicit specification of a conceptualization" [74]. In artificial intelligence, the term ontology refers to the concepts, definitions, and relationships that make up a model to support semantic processing [75]. A formal ontology is a controlled vocabulary expressed in an ontology representation language. It

includes concepts, instances, attributes, assertions, and logical constraints of those concepts, instances and relationships [76-77].

Each of these discussed terminology structures are approaches to help structure, classify, model, and or represent the concepts and relationships in a specified domain. Each structure provides a resource for establishing community agreement, commit to term use, and foster discussion on appropriate definition and usage. Other terminology notable standards such as those in Chute et al.'s framework for comprehensive health terminology systems in the US [69] and the International Standards Organization's technical specifications [78] discussed guiding principles to be considered when building terminologies regardless of format or technology. If a taxonomy has a variety of very carefully defined meanings for the hierarchical link, then it bears a stronger resemblance to an ontology [77-79].

2.4 Medical Terminology Resources

There are many available terminology resources in the biomedical and life science domain. Medical terminologies rapidly grew in popularity over the last couple of decades as a method for representing events and healthcare data. Development and acceptance was motivated because these terminologies supported reimbursement and regulatory compliance [21]. Terms cover diseases, diagnoses, findings, operations, treatments, drugs, administrative items etc., and can be used to support reporting both research findings and patient care at varying levels of detail. A nomenclature is a relatively simple system of names. A vocabulary is a system of names with explanations of their meanings. A classification is a systematic organization of things into classes, and a thesaurus is designed to index medical literature and support search over bibliographic databases. But many of the terms used in this field can prove difficult to define accurately, and their use in practice can be inconsistent. As stated by Elkin, "the goal of healthcare terminologies was and is to aggregate patient descriptions by meaning" which require reduced ambiguity in an iterative development environment. The linguistics or rules of language provide important techniques for decoding the medical terminology [21].

An esteemed medical concept terminology with successful application support implementation is the General Architecture for Languages, Encyclopedias and Nomenclatures in Medicine (GALEN) [80]. Developed with a concern for computerization of clinical terminologies, it has assisted in the referencing of electronic medical record systems, decision

support and other clinical systems. It was one of the first architectures to explore the implementation of description logic for medical applications [81]. Its aim is not to be a comprehensive resource but to make available a sufficient amount of concepts in order to build upon OpenGALEN (www.openGALEN.org), a distributable reference model to support applications [80, 82].

The National Library of Medicine (NLM) maintains the Unified Medical Language System (UMLS) [83] which is a comprehensive collection of over 100 terminologies, 1 million concepts, and 4 million strings for those concepts. It provides knowledge sources and tools for facilitating the development of software applications referencing the terminology of biology, medicine, and health [84]. As stated by Saripalle, the UMLS has probably a greater impact on biomedical ontology work than any other terminology effort because of its long history, its early focus on knowledge representation and its size and free availability [85]. It, as a system, supports syntactic analysis with several lexical tools and contains a set of computer programs that process natural language words and terms including free text narratives [21]. It consists of three components that provide structured knowledge in the biomedical domain: the SPECIALIST Lexicon [86], the Semantic Network [87], and the Metathesaurus [88]. The SPECIALIST Lexicon [86] is a resource to support natural language processing (NLP) describing syntactic characteristics, part-of-speech labels, and inflections. The Metathesaurus [88] is a multilingual vocabulary database that allows biomedical and health-related concepts, their various synonyms, and the hierarchical information to be identified in text. Terms from each source vocabulary in its database are organized by meaning and assigned a concept unique identifier (CUI) [89]. Each concept in the Metathesaurus is assigned at least one semantic type from the Semantic Network which constitutes a form of an upper-level ontology. Semantic types are mostly broad subject matter categorizations and are useful linkages that exist between different semantic types [21, 87].

The Medical Subject Headings (MeSH) is a biomedical and health science controlled vocabulary maintained by the NLM for PubMed. PubMed is the search engine for NLM's bibliographic database, MEDLINE, which includes records from millions of biomedical and health science journal articles [90]. MeSH alphabetically and hierarchically organizes medical knowledge and information into a thesaurus [91].

One of the most notable sub-domain knowledge bases is the Gene Ontology (GO) created by a consortium of molecular biologists. proteomic and genomic bioinfomaticians [92]. The Gene Ontology Consortium is made up of many research groups synthesizing several highly specialized biological databases [93]. The GO was developed with principles described by the Open Biomedical Ontologies (OBO) for developing controlled vocabularies for shared use across various biomedical domains. Goals of the Gene Ontology Consortium focuses on evolving knowledge of genes' and proteins' roles in cells aggregating the largest collection of molecular biology-related terminological representation. In-depth coverage of cellular components, molecular functions, and biological processes is represented and updated by members of the research and annotation communities, as well as by those directly involved in the GO project [94]. The database integrates vocabularies and annotations for access in several formats interoperable with many functional applications. This project has been an example of the power in community collaboration to develop standard terminology and achieve acceptable levels of definitional agreement and conceptual logic [95-96].

Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) is a comprehensive multilingual clinical healthcare controlled vocabulary which is clinically validated and mapped to other international standards [84, 97]. SNOMED-CT has a long history of aggregating medical terminology lists, hierarchical representation of concepts, and expressional definitional knowledge of terms [98]. Its purpose is for the exchange of clinical health information. Clinical findings, symptoms, diagnoses, procedures, body structures, organisms and other etiologies, substances, pharmaceuticals, devices and specimens is included in its coverage. SNOMED-CT contains more than 300,000 medical concepts represented by an individual number and structured where several concepts can be used simultaneously to describe a complex condition reducing ambiguity from the use of regional or colloquial terms [94, 99].

The US National Cancer Institute (NCI) developed the NCI Thesaurus (NCIT) to support cancer research based on current biomedical science, such as diseases and underlying biology [100]. The NCI Thesaurus provides resources and services to meet the NCI's needs for controlled terminology and to facilitate the standardization of terminology and information systems across the institute and also the biomedical research community [101]. It covers terminology for clinical care, translational and basic research, and public information and administrative activities [102]. There are more than 400,000 relationships between concepts to

help define and connect them. The concepts include codes, terms, abbreviations, synonyms, definitions, links to outside sources, and additional supportive information [103].

The Logical Observations, Identifiers, Names and Codes (LOINC) is a terminology that represents laboratory and other diagnostic tests [104]. The categories include chemistry, hematology, serology, microbiology, toxicology and drug categories. It was originally developed at Regenstrief Institute of Indiana University, however the LOINC community has expanded around the globe to provided coverage to additional clinical domains. This clinical coverage includes vital signs, hemodynamics, EKG findings, echo findings, urologic imaging findings, and pulmonary ventilator management to name a few [105]. LOINC also provides a standard for coding observations in HL7 messages to promote exchange of electronic health information [104]. It is a terminology where each concept includes a fully specified name and widely adopted standard for the medical laboratory community. LOINC has facilitated transfer of public health reporting, reduced errors, and supported aggregation of EHR data [105-106].

The International Classification of Diseases (ICD) supports general purpose disease classification congruent with the World Health Organization (WHO) classifications. This system provides a common language framework for governments, providers, and consumers to share information [107]. There are several versions of ICD available but each facilitate the storage, retrieval, analysis, and interpretation of data. ICD-9 was released in 1979 and is still in use today, for collecting and classifying mortality statistics. While ICD-11 is in development, ICD-10 is the current version with an additional volume set, alphanumeric categories over numeric, category rearrangement, and recoding of rules for mortality [108]. It reports causes of death translating them into medical codes while consolidating conditions and incorporating multiple causes of death [109-110].

The American Medical Association developed and maintains the Current Procedural Terminology (CPT) code set. In the United States, CPT is virtually used by all public and private healthcare payers, all healthcare professionals, and institutional providers [111]. Primarily used to report services and procedures reported on health insurance claims, the first edition of CPT was published in 1966 [112]. The current edition includes numerical codes with descriptors for reporting medical services and procedures performed by physicians and other healthcare professionals [113]. In the context of reporting services, CPT provides a consistent language to describe medical, surgical, and diagnostic services [111, 114].

There are also several nursing vocabularies that are very thorough and foster usability in terminology system framework design. Development in nursing terminology has been successful, exemplified by Nursing Diagnoses, Definitions, and Classification (NANDA) [115-116], Nursing Interventions Classification System (NIC), [117] Clinical Care Classification (CCC), Nursing Outcomes Classification (NOC) [118], and the International Classification for Nursing Practice (ICNP) [119-121].

2.5 Mental Health Terminology Resources

Mental health terminology resources support the development of clinical and consumer applications, decision-making, and foster communication among stakeholders of PTSD and various other disorders. Terms utilized in the mental health domain extend their dictionary definitions deeper and are, by nature, embedded with a greater amount of assumption [122]. Normal frames of reference can vary in the mental health world and even between its overlapping disorders and diagnoses. These terms are imperative to codify objects, make assumptions explicit, and eliminate false perceptions. The descriptors that make up these mental health terminology resources must be inclusive and flexible as information grows, changes and understanding of these disorders improves. Terminology systems contribute to the requirements needed for data management and computational support. This focus is needed because of the lack of objective measures for the disorder whose primary validity bases diagnoses on consensus clinical symptom clusters [123]. The collective mental health terminologies must provide reliable recognition of each disorder with consistent characterization, permit accurate diagnosis from multiple data points, and inform treatment options.

The International Classification of Diseases (ICD-10) produced by the World Health Organization (WHO) [108] and the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) [6] produced by the American Psychiatric Association (APA) are the two widely established systems for classifying mental disorders [123]. Both are broadly comparable via lists of disorder categories thought to be distinctive types, which have converged from revisions and local installation alignments [122].

Published International Literature on Traumatic Stress (PILOTS) [124] is a bibliographic database maintained by the Veterans Administration's National Center for PTSD [125]. It is supported by a thorough and consistent thesaurus in order to guide a diverse set of users through

an equally diverse amount of literature. The PILOTS Thesaurus is essentially a purpose-built controlled vocabulary used to index and retrieve literature [49]. The database does not restrict its coverage to specified journal articles but instead strives to capture information in all forms. The database contains articles, books, posters, gray literature, pamphlets, white papers, and all materials of practical value for issues surrounding the disorder [126]. The complete database of PILOTS is searchable via the ProQuest platform with basic and advanced search features in order to manage all forms of traumatic stress from worldwide resources [125].

Hadzic et al. developed the mental health ontology (MHO) [127] as a resource to support automated tasks with text and concepts. Mental illness, its various causes, and treatment information is collected and managed as three substructures which represent (1) disorder types, (2) factors and (3) treatments [127]. These substructures were modeled with the ICD-10 classification system fostering interoperability with existing systems.

The mental functioning ontology (MF) [14] is maintained to capture many features in all areas of mental functioning. This includes mental processes of thinking, planning, learning, remembering, reasoning, and intelligence qualities. It is developed according to OBO Foundry [128] guidelines using the Basic Formal Ontology (BFO) [129] as its upper-level ontology. MF uses modular creation steps, allowing extensions to be created in sub-domains of mental functioning or related areas where terminology creation is needed [130]. In coordination with the mental functioning ontology, the mental disease ontology (MD) is created modularly to categorize mental disorders based upon the work of Ceusters and Smith [131]. It incorporates BFO and Ontology for General Medical Science (OGMS) for integrating clinical research and data from EHRs. OGMSs also supports interoperability with other related terminology resources, increasing the coverage of MD and its impact [130].

Khoozani et al. provides a hierarchical model to represent the human stress ontology (HSO), aggregating knowledge researchers have accrued. It consists of stress causes, stress mediators, stress effects, stress treatments, and stress measurements. HSO is beginning development towards supporting application development, and assisting researchers and clinicians engaged in treating the effects of stress on the human body [132].

Created as an extension to the mental functioning ontology, the emotion ontology (MFOEM) has a strong focus on emotions and moods separating them from symptomatology. Built upon the BFO, it is described as a collection of “phenomena such as emotions, moods, appraisals and

feelings" [133] all of which are highly subjective. MFOEM supports research by providing unified annotations building upon the coverage of the mental functioning ontology [130-131].

2.6 Terminology Development

Terminology engineering begins with requirements gathering followed by thorough knowledge acquisition to achieve breadth of coverage that is necessary to ensure extensive usage. Many terminology development methodologies follow features outlined by Uschold and King [134] for ontology engineering. In summary, the steps are: 1) identify purpose and scope; 2) build, code, and integrate existing resources; 3) evaluate; and 4) document. Critical for long-term success of a developed terminology system for text mining is putting it to use in the annotation of data. The terms must be identified, linked to meaningful concepts, and mapped to available existing terminology resources [21, 135-136].

Requirements gathering describes the content, functionality and quality necessitated by stakeholders. For the identification of user needs, the requirements specify the goals and tasks the system should perform. Tasks, activities, constraints, and preferences are determined by requirement recognition. The requirements themselves are the descriptions of the system services and constraints that are generated during the engineering process. Importantly, stakeholders' wants are specified rather than the focus of how they will be delivered. The categorization of users is defined as well as their characteristics such as prior knowledge and experiences, special needs, and subjective preferences [13, 21].

In the process of identification and analysis, user surveys, use case scenarios, or interviews are carried out. Use cases provide detailed realistic examples of tasks in a specified context. These are compiled to provide a vivid representation of the envisioned use of the system. Questionnaires can be distributed to a sample population of users to determine needs and current workflow considerations. Users are interviewed, in most cases, based on a predefined questionnaire. Although the interviews are carried out based on a series of fixed questions, users should be prompted to elaborate on their responses as well other related interesting issues that may not be included in the interview questionnaire [12-13].

Terminology models are necessary in order to share and re-use knowledge, but various approaches can be used to collect this domain knowledge [137]. Development of the PTSD terminology framework focuses on identifying content and terms that are specific and exhaustive

to the PTSD sub-domain. This terminology framework is constructed in the form of a controlled vocabulary incorporating a hierarchical classification structure. In the framework, a concept can be referred to by more than one term and each of these terms can be expressed at different levels of generality. The development of controlled vocabularies and taxonomies can be considered one of the early stages of developing a domain-specific ontology [21, 138].

In addition to the agile methodologies described in section 3.2.4, the majority of both the requirement and terminology framework development for this project is derived from several validated ontology development methods listed in **Table 2.2**. Terminology framework requirements account for end-users but places more emphasis on the additional stakeholders and their respected needs that are identified in the engineering process. Developments in areas such as information retrieval, artificial intelligence, and human computer interaction, are often pertinent and mirror techniques and methods of software and database requirements engineering [138-140]. The vocabulary for this research was geared toward being built first as a generic view of the PTSD sub-domain of medicine from which more specific, task-oriented iterations are implemented. The advantage of this strategy to the terminology, taxonomy, and ontology community will align consistent methods contributing to further integration and reuse between ontologies and vocabulary resources [78, 141-142].

Method
• Ontology development 101 • Neon methodology • Grüninger and Fox method • Uschold and Grüninger method • Methontology

Table 2.2. Applicable ontology development methods [75, 138-142]

There are numerous strategies for efficient knowledge acquisition for terminology development. Linguistic techniques utilize automatic and semi-automatic extraction from text based on corpus-driven concepts and relations. NLP-driven templates support acquisition via rule-based text extraction to identify needed concepts. Lastly, there are hundreds of existing ontologies and terminology resources to extend domain-specific knowledge acquisition [143-144].

2.7 Biomedical Text Mining

Text mining uses information extraction and is defined by Hearst [145] as the "process of discovering and extracting knowledge" from unstructured data. Text mining typically comprises of information retrieval, information extraction, or data mining. Information retrieval aims to gather relevant texts. Information extraction extracts specific types of information from texts of interest. Data mining finds associations among the extracted pieces of information [146]. According to Hersh, text mining utilizes information retrieval, data mining, machine learning, statistics, and computational linguistics in order to deriving information from text. These methods are utilized to find patterns and trends in the textual information [16].

As a component of artificial intelligence, natural language processing (NLP) fosters software applications to understand human speech as it is spoken. According to Bird et al., it implements "algorithms that allows human languages to be manipulated by computers to provide structure to data" [147]. Meystre et al. has recognized that because of medical data growth in all types of formats, the incentives for developing of NLP systems is expanding as well [146]. Common NLP tasks include sentence segmentation, part-of-speech tagging, parsing, word sense disambiguation, deep analytics, co-reference resolution, information retrieval, information extraction, and named entity extraction. NLP methodologies in the biomedical domain can be considered from the point of view of the text they address and the NLP technology used. Two important content subdomains are clinical medicine and molecular biology [84]. Linguistic approaches are often categorized broadly as symbolic rule-based or statistical systems. Due to the complexity of language, systems often focus on one aspect of linguistic structure: words, phrases, semantic concepts, or semantic relations. Words can be identified with little minimal linguistic processing [148]. Phrases are normally identified on the basis of at least some syntactic analysis, using part-of-speech categories and rules for defining phrase patterns [149]. The identification of concepts and relations constitutes semantic processing and requires that text be mapped to a knowledge structure. In the biomedical domain, the UMLS provides one such resource, however its coverage is incomplete for PTSD as well as other mental health disorders [84].

The field of information extraction (IE) began with the Message Understanding Conference (MUC) [150], which consisted of scientific and shared specialized domain tasks. Participants in these tasks had to answer specific questions concerning several special mentions in the text.

Answering the evaluation set of questions automatically is a non-trivial task. Because of that, many systems were handcrafted and contained a huge amount of manually designed rules, resulting in systems biased towards a specific domain. Using these systems for other domains was impossible or very unattractive, because much human effort had to be invested to tune all the rules manually.

IE typically requires some pre-processing such as spell checking, document structure analysis, sentence splitting, tokenization, word sense disambiguation, part-of-speech tagging, and some form of parsing [146]. A variety of methods have been employed in the general and biomedical literature domains to extract facts from free text and fill out template slots as described by McNaught et al. [151]. An IE system often consists of a combination of the following components listed by Hobbs [152-153]: tokenizer, sentence boundary detector, part-of-speech tagger, morphological analyzer, shallow parser, deep parser, gazetteer, named entity recognizer, discourse module, template extractor, and template combiner. The information extracted can then be linked to concepts in standard terminologies and used for coding [146]. This can be accomplished by pattern-matching over text strings, part-of-speech tags, semantic pairs, and dictionary entries [154]. There have also been initiatives at ontology-driven IE to guide the free-text processing [155]. Machine learning techniques have demonstrated remarkable results in the general domain and hold promise for clinical IE, but they require large, annotated corpora for training, which are both expensive and time-consuming to generate.

The Linguistic String Project-Medical Language Processor (LSP-MLP) [156] is an early extraction and summarization IE system developed by Sagar that examined signs/symptoms, drug information, and identification of possible medication side effects. Friedman at al. [157] expanded upon Sager's efforts to develop the MedLEE (Medical Language Extraction and Encoding system) system to support natural language queries by extracting information from clinical narrative reports [158]. Another early system was SPRUS (Special Purpose Radiology Understanding System) [159] which was a semantically driven NLP application. Also noteworthy, is SymText (Symbolic Text processor) [160] which relied upon Bayesian networks for syntactic and probabilistic semantic analysis. The U.S. National Library of Medicine has developed a set of NLP applications called the SPECIALIST system [161], as part of the UMLS project [162]. It includes the SPECIALIST Lexicon, the Semantic Network, and the UMLS

Metathesaurus. The NLM also developed several applications that use the UMLS, such as the Lexical Tools and MetaMap [88] that provided many features of a complete IE system.

Named entity recognition (NER) is a sub-field of information extraction and refers to the task of recognizing expressions denoting named entities in databases or free text documents [163]. It is also known as entity identification, entity chunking and entity extraction. It seeks to locate and classify elements in text into pre-defined categories. The mapping between terms and concepts in a terminology resource is not inconsequential [164]. The entities in this research are predefined categories of symptoms and therapeutic interventions to be identified. NER systems have been created using linguistic grammar-based techniques as well as statistical models and machine learning approaches to locate, then classify entities [16]. A lexicon-based approach relies primarily on the quality of its referential terminology source although a certain amount of rules within the system is typically involved. Rule-based NER systems can be very effective, but require some manual effort. Machine learning approaches can successfully extract named entities but require large annotated training corpora. Advantages of machine learning approaches are that they do not require human intuition and can be retrained without reprogramming for any domain. However, alterations to program code can be vastly difficult to maintain if proper change management and documentation is not sustained [165]. NER is in no way domain-independent, because every special domain needs special entities to be extracted.

There are several successful implementations of NER systems. Meystre et al. review information extraction from the clinical narrative [146]. A very popular application area of NER is the biological domain which uses NER for identifying genes and proteins. This task is referred to as bio-entity recognition and example research was demonstrated in the BioNLP of 2004 [166]. Liu et al. studied the comprehensiveness of the UMLS and found it to be an “invaluable source” for NLP but lacked coverage to support biomedical applications for clinical text. The authors found that using it as a foundational source was useful to expand a semantic lexicon for semantic parsing [167]. An excellent synopsis of dictionary-based, rule-based, statistical, and hybrid approaches to automatic named entity recognition for biomedical literature is provided in Krauthammer and Nenadic [165]. Chiang and Yu [168] used a rule-based approach and the Gene Ontology to support dictionary-based term recognition. Variants arising from word order variations as well as missing or additive word tokens were considered when selecting an entity. A token is an instance of a sequence of characters in some particular document that are grouped

together as a useful semantic unit for processing. The authors also calculated edit distance as a way of quantifying string dissimilarity to measure term variant recognition

2.8 Natural Language Processing Frameworks

Natural language processing (NLP) frameworks provide the environment to refine NLP tasks and improve the output of data processing. The most successful frameworks are module-based and managed as compatible objects for a variety of implementations. A typical architecture is object-based supporting a variety of data structures for analysis. Common artifacts are text documents, but they can be other things, such as video and audio streams [18].

NLP frameworks support integration with search technologies and analysis of unstructured information. Analysis components will typically include tokenizers, summarizers, categorizers, parsers, and named-entity detectors etc. Applications are developed to detect, organize and interpret relevant information not already explicitly annotated. The goal is to provide structure to the unstructured information for end-user data processing or queries. Frameworks use a variety of technologies including information retrieval, terminology resources, statistical and rule-based algorithms, automated reasoning, and machine learning. The frameworks provide the essential components for developing specific applications using varying interfaces and programming languages [169-170].

2.8.1 Selected NLP Frameworks

General architecture for text engineering (GATE) is a suite of Java tools for performing NLP and IE tasks. It provides a common infrastructure supporting many languages, machine learning plugins, and various formats of textual input. It also provides a testing and evaluation environment. GATE consists of a database and a database schema, a graphical interface, and a collection of wrappers for database interoperability. Also included are machine learning plugins for Weka, RASP, MAXENT, and SVM Light [171-172].

The natural language toolkit (NLTK) is a Python-based suite of libraries for symbolic and statistical NLP tasks. It has the perception of being a text mining learning toolkit due to the fact that a majority of its text processing capabilities can be performed Python's basic data structures. However, many of its components can be modified for completing more complex NLP tasks.

The visualization modules are excellent viewing and manipulating data analysis experiments [147, 173].

The Apache OpenNLP is a Java-based machine learning toolkit for performing NLP tasks. Its library includes rule based and statistical NLP. It is a mature toolkit containing a library of several components for developing a complete NLP pipeline. It also includes a wrapper for integration with other frameworks for performing automated annotation and training new OpenNLP models from annotated text [174].

Written in Java, Apache Lucene [175] is an open-source information retrieval software library. The logical architecture organizes each document as a collection of fields irrespective of its original file format. Lucene indexes can be implemented to search fields of text within documents and supports many linguistic operations such as tokenization and stemming in various programming languages. Regarded as a utility of search engines, Lucene indexes can support many applications that need full text indexing and searching [175]. One of the major appeals of this software library is its ability to search multiple file formats containing extractable text. Using Lucene is essentially a two-step process: 1) create the index on documents or database, and 2) parse a query by utilizing the prebuilt table [176]. However, the configuration options are extensive providing query ranking where developers can apply domain-specific features to text improving the relevance of results [175].

The machine learning for language toolkit (MALLET) is a Java-based statistical NLP platform capable of machine learning tasks, document classification, cluster analysis, information extraction. The toolkit implements a variety of algorithms for text analysis and supports several add-ons for generating graphical models [177].

Deep linguistic processing with HPSG - initiative (DELPH-IN) [178] is a computational linguist collaboration building NLP tools by applying head-driven phrase structure grammar (HPSG). HPSG is a lexical non-derivational generative grammar theory based upon a system of rules to form sentences from prior word combinations. DELPH-IN framework provides module pipeline engine system development architecture capable of supporting rule and statistical-based text mining [179].

Stanford CoreNLP (http://nlp.stanford.edu/software/) provides a set of natural language analysis tools including a Java-based conditional random fields and named entity recognition tool. Their language technology is designed to sentence understanding, machine translation,

probabilistic parsing and tagging, biomedical information extraction, grammar induction, word sense disambiguation, and building question answering systems. Its analyses provide the fundamental building blocks for developing higher-level and domain-specific text understanding applications. Stanford CoreNLP includes several grammatical analysis toolkits and supports various programming languages interfaces for pipeline module conception [170].

Unstructured information management architecture (UIMA) is a self-described "industry standard for content analytics" (https://uima.apache.org/) [169]. It is a framework for performing NLP tasks to structured various forms of unstructured information such as text, audio, video, images, etc. UIMA enables applications to be decomposed into components where each component implements interfaces defined by the framework and provides self-describing metadata via XML descriptor files. The framework manages these components and the data flow between them. Components are written in Java or C++, and the data flow between components is designed for efficient mapping between these languages [169, 180].

2.8.2 Text Processing System Pipelines

The Linguistic String Project (LSP) is one of the earliest initiatives at an NLP system to implement a parsing program as a first step in the computer processing of the scientific literature. The original goal was to answer queries enabled by the retrieval of specific information for investigators. The LSP sought to automate the application of health care standards to information formatted narrative medical reports. It concentrated specifically upon X-ray reports, hospital discharge summaries, and the sublanguage of clinical reporting [181]. The system developed by the project came to represent its own programming language developed for clinical narrative. The programs, developed by the project adapted for clinical narrative in LSP Medical Language Processing (LSP-MLP) that supported online access by clinicians to portions of narrative patient documents relevant to stated concerns [182].

Using a controlled vocabulary, Medical Language Extraction and Encoding System (MedLEE) is an NLP system that extracts information from clinical narratives [183]. It uses a lexicon to present information in a structured form by mapping terms to classes and semantic grammar. Appropriate controlled vocabulary support is critical to the success of MedLEE implementations and output is based upon XML allowing further incorporation of localized terminology [184]. It is primarily configured for application to medical reports, discharge

summaries, pathology reports, and radiology reports [185]. MedLEE has been successful at extracting information from the textual narrative to support decision making, clinical event screening, and safety improvement through intervention [186].

Health Information Text Extraction (HiTEx) is a rule-based NLP tool developed with the GATE framework. With HiTEx, module-based text mining pipelines can be created to extract specified findings from clinical narrative text. It has primarily been implemented for several low-level NLP mining tasks and has a module for implementing a UMLS concept mapper. Sequential module tasks such as section splitters, section filters, sentence splitting, sentence tokenizers, POS taggers, noun phrase finders, UMLS concept mapper, and negation finders have been successful at physician diagnosis identification for narrative free text of medical reports [187].

RapidMiner is a Java-based software platform for data mining and analytics. The appeal of this platform is the speed and ease of implementing data mining and machine learning techniques via template frameworks [188]. RapidMiner provides a graphical user interface for development and allows extensions via R and Python scripts. Specifically, this platform has been successful at IE and NER implementations applicable to this PTSD text mining research [188-189].

MetaMap [88] is a configurable program developed at the National Library of Medicine (NLM) in order to map biomedical text to the UMLS Metathesaurus. It has also applied linguistic principles to discover Metathesaurus concepts referred to in text. MetaMap applies advanced computational linguistics, statistical methods, and symbolic NLP for application to information retrieval and data mining. The UMLS Metathesaurus forms the core of the UMLS and incorporates over 100 source vocabularies to find mentions of clinical terms based on CUI mappings [190]. The program links text primarily in published literature to biomedical knowledge including synonyms within the Metathesaurus. After low-level NLP tasks are performed, words are identified by dictionary lookup in the SPECIALIST lexicon and shallow parsing using the SPECIALIST parser [191]. Mapping to UMLS concepts is performed followed by word sense disambiguation to identify a suggested concept identification, mapping construction, and its candidate evaluation [190-192].

The NCBO Annotator [193] is an online system maintained by the National Center for Biomedical Ontology (NCBO) that utilizes over 300 ontologies in BioPortal for identifying concepts in unstructured text. The system performs concept recognition by lexical matching of terms and their synonyms followed by processing via mgrep, a multi-line grep tool to produce

annotations [194]. Though mixed results are reported, NCBO Annotator provides an automated alternative form to manual annotation in order to train algorithms. The NCBO Annotator has shown good performance on concept annotations when compared with MetaMap, however users often report large system downtime and dissimilar results due to constant changes to user community ontologies. A representational state transfer (REST) Web service from NCBO permits application implementation outside of the center's programming environment [192, 195].

SemRep [196] is an NLP tool developed for the biomedical research literature to provide semantic interpretation of text supported by domain knowledge of the UMLS [197]. The system is symbolic and rule based directed by linguistic techniques described in Rosemblat et al. [198]. Processing begins with a lexical analysis based on the SPECIALIST Lexicon stochastic tagger [86]. Textual content is assigned via Metathesaurus concepts and semantic types using MetaMap [88], followed by predicates from relations in the Semantic Network [87]. For example, from the text in (1), SemRep identifies the concepts and their corresponding relationship in (2).

(1) **Sertraline** was used in the treatment of **PTSD** patients

(2) Sertraline **TREATS** PTSD

SemRep identifies: a. Metathesaurus concept of sertraline (semantic type: Pharmacologic Substance); b. Stress Disorders, Post-Traumatic (semantic type: Disease or Syndrome); and c. the predicate 'treats' via the Semantic Network. SemRep has been used in literature-based discovery (LBD) [199] and for creating executable knowledge for information management [84]. Rosemblat et al. [198] has developed methodology to add enhancements for improved domain coverage beyond clinical medicine and basic biomedical research.

The clinical Text Analysis and Knowledge Extraction System [200] (cTAKES) is a statistical and rule-based text-mining pipeline. NLP tasks are module-based and comes supported with OpenNLP boundary detection, tagging, parsing, tokenization, normalization via SPECIALIST lexical tools, and negation via NegEx [201]. The cumulative sequential execution of these modules produces a complete annotated dataset. The features of the annotation prepare the data for even more advanced clinical semantic processing. cTAKES accepts input of XML-based continuance of care document as well as plain text [192]. Each named entity recognized with the cTAKES NER component maps to a concept from the terminology linked to the specified

module. One important shortcoming of this system is that ambiguities that manifest via multiple terms within the same span of text cannot be resolved [200]. Ambiguities can be addressed with specified dictionary implementation via additional programming of modules implemented into the NLP pipeline. Projects such as the Unified Medical Language System (UMLS) [162] and the Linguistic String Project (LSP) [181-182] has a history in addressing this ambiguity with this approach by developing schemas of clinical text and a dictionaries of medical terms to support NLP [202]. Authors such as Resnick [203] has addressed resolving this syntactic and semantic uncertainty with adncened algorithms implemented with NLP pipelines. Specifically, in the mental health domain, Hadzic et al. [204] implemented data mining techniques for analyzing semi-structured mental health data with great efficiency. Fodeh et al. [205] developed a framework to reduce dimensionality and noise in PTSD clinical notes by a concept-driven approach for improving understanding and summarization.

2.9 Data Mining

Data mining is the practice of applying techniques to search large amounts of computerized data to find useful patterns or trends [202]. Patterns and correlations are found using sophisticated algorithms for data processing. The identification of salient information is difficult because of the challenges of acquiring and representing medical knowledge. For example, drug development costs were decreased by Epstein [203] by identifying unknown relations by text-based data mining of scientific literature to refine therapeutic hypotheses. In the behavioral health domain, Panagiotakopoulos et al. mined the treatment of anxiety disorders via data collected for long-term monitoring by developing a contextual data mining approach [204]. Continued advancement in data mining technology will foster the development of knowledge bases of information for determining the necessary patterns and trends. Along with comprehensive terminology structures and thorough NLP pipelines, the future of accurate clinical decision support systems will require outstanding data mining techniques [18, 21, 205].

2.9.1 Selected Data Mining Algorithms

First used by Tryon [206] in 1939, the term cluster analysis encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. Unlike many other statistical procedures, cluster analysis methods are mostly used when an a

priori hypotheses is not available, and researchers are in the exploratory phase. There are over 100 published clustering algorithms and this data mining technique has been useful to healthcare researchers in many contexts. In 1975, Hartigan discussed a thorough set of summaries of the many published studies reporting the results of cluster analyses. For example, in the field of medicine, clustering diseases, cures for diseases, or symptoms of diseases can lead to very useful taxonomies. In the field of psychiatry, the correct diagnosis of clusters of symptoms such as PTSD, paranoia, schizophrenia, etc. is essential for successful therapy [207-208].

Different approaches to clustering data can be described from the listing shown in **Table 2.3**. Hierarchical clustering falls under a connectivity model for building a hierarchy of clusters. K-means aims to partition *n* observations into *k* clusters by finding the closest similar centroid. In k-means, each observation belongs to the cluster with the nearest similar mean. Expectation maximization clustering performs a function for maximizing the expectation of the log-likelihood [207]. A technique introduced by Mirkin [209], biclustering or co-clustering allows simultaneous clustering of the rows and columns of a matrix.

Algorithms	Examples
Connectivity models	hierarchical clustering, UPGMA linkage clustering
Centroid models	k-means
Distribution models	expectation maximization (EM)
Density models	DBSCAN, OPTICS
Subspace models	Biclustering

Table 2.3. Clustering algorithms available for implementation [207]

A cluster is a collection of data objects that are similar to one another. The similarity can be measured in many ways such as lexical distance, semantic meaning, or co-occurrence probability matrix distance. In a cluster analysis, the objects are grouped based on this defined similarity between each data class and maximized difference between other class groupings [210-211]. Term clustering is the grouping of similar words, based on their tendency to co-occur in similar contexts [212]. It was introduced by Brown et al. [213] and used in different applications, including named entity tagging, machine translation, and text categorization. In most of the studies in term clustering, co-occurrences appearing in the same document, in the same sentence or following the same word has been used to estimate term similarity [212]. Prior research has

approached problems of clustering words based upon co-occurrence data, and using the acquired word classes to improve the accuracy of syntactic disambiguation [214]. This research utilized co-occurrence of concepts as features for machine learning similar to the work of Zhang et al. where co-occurrence features of geo-temporal distributions of tags were extracted and represented as vectors for clustering [215]. According to Jain et al., feature selection is the "process of identifying the most effective subset of the original features to use in clustering" [207]. The feature utilized in this research is co-occurrences of concepts. This similarity-based clustering of word co-occurrence probabilities is thoroughly described by Cardie in 1993, Ng in 1997, and Zavrel et al. in 1997 [216-218].

Discovery of association rules is an important component of data mining. Association rule learning can find patterns in data which reveal combinations of events that occur at the same time. Association rules have wide applicability and have been useful in many areas of nuclear science, pharmacoepidemiology, immunology, bioinformatics, and healthcare [219]. In association mining, the emphasis is almost always on large lifts or positive associations. The two main applications of association mining are market basket analysis and finding prediction rules. In market basket analysis, the dataset consists of a collection of sets baskets and are used to find elements that frequently co-occur together in these sets. Recent studies have shown that there are various algorithms for finding association rules, but one of the best known is the apriori algorithm [219]. An association analysis identifies correlations that occur frequently together among a set of items. Classification utilizes training data to distinguish between concepts to find patterns within the data. It exploits a model to make predictions about the classes of the objects it explores [207].

Association rule mining finds interesting associations and/or correlation relationships among large sets of data items that occur frequently together in a given dataset [219]. The definitions for the terms utilized in the association rule mining algorithm for gathering and evaluating are shown in **Table 2.4**. Overall, lift summarizes the strength of association between the products on the left and right hand side of the rule; the larger the lift the greater the link between the two products [219]. An important characteristic of association rule mining is that it divides the problem of mining into sub-problems to do efficient computing [220].

Term	Definition
Transaction	Identified with a unique identifier, it is a set of items with a minimum of one item.
Items	Depending on the application field, they can be products, objects, patients, events.
Itemset	A group of items. Itemsets can be found in one or more transactions.
Rule	A rule defines a relationship between two itemsets X and Y that have no items in common. X->Y means that if we have X in a transaction, then we can have Y in the same transaction.
Support	The probability to find items or itemsets in a transaction. The higher the support the more frequently the itemset occurs. It is estimated by the number of times the items are found across all available transactions.
Confidence	The probability that a transaction that contains the items on the left hand side of the rule also contains the item on the right hand side. The higher the confidence, the greater the likelihood that the item on the right hand side will occurr.
Lift	The ratio of confidence to expected confidence. Lift is a value that gives us information about the increase in probability of the "then" (consequent) given the "if" (antecedent) part.

Table 2.4. Definitions for association rule mining [219]

Nuwangi et al. [221] researched the global prevalence of diabetes mellitus with the apriori algorithm using the Waikato environment for knowledge analysis (Weka) [222]. Association rule mining found multiple complications of diabetes considering gender, age and occupation factors. This research produces some new results that were earlier not given due weightage but are significantly important in the medical field [220]. Witten et al. developed a method using apriori and implemented the algorithm on a large medical dataset using the proposed technique [202] in Weka. Lekha et al. presented a new method [223] to generate association rules on numeric data using apriori algorithm and classification technique on a diabetes dataset [220].

2.9.2 Selected Machine Learning Toolkits

Weka [222] is a Java-based suite of machine learning toolkits often implemented for data analysis and data mining applications. It provides graphical user interfaces for implementing visualization tools and algorithms. Weka provides a collection of data preprocessing and

modeling techniques and access to SQL databases using Java Database Connectivity (JDBC) [224-225].

Scikit-learn is a Python-based machine learning library designed to interoperate with the widely popular libraries NumPy and SciPy [226]. NumPy supports large, multi-dimensional arrays and matrices, along with a library of mathematical functions to operate on these arrays [227]. SciPy builds on the NumPy array object containing modules for optimization, linear algebra, integration, interpolation, special functions, fast fourier transform (FFT), signal and image processing tasks [228]. Scikit-learn provides both supervised and unsupervised machine learning algorithms in a framework for interoperability into applications outside the domain of traditional statistical analysis software. It incorporates the C++ libraries LibSVM [229] and LibLinear [230] that provide reference implementations of SVMs, and other generalized linear models [226, 230].

The Microsoft SQL data mining and predictive modeling plugin [231] requires installation and processing on a server to perform business intelligence on specified data. The proficiency behind the data mining plugin resides in the processing power of the SQL Server components of the database engine, integration and analysis services. The analysis services implement raw data examination using Online Analytical Processing (OLAP) cubes and data mining algorithms. Clustering algorithms detect categories, association rule mining discovers data relationship correspondence, and built-in tools identify anomalies in the data. Decision trees, neural networking, naïve Bayes, linear regression, and logistic regression are all special cases of the Classification and Regression Tree (CART) algorithms to quickly build data mining models [232-233].

2.10 Gaps in Literature and Knowledge

The range of symptoms, variability of patient reaction to emotions, prevalence of misinformation, and contradicting treatments all contribute to complicate an overall understanding of PTSD [8]. Despite large monetary efforts to combat the disorder, there lacks a consensus definition of symptom manifestation exhibited among patients. Additionally, there is constantly changing parameters for measuring resiliency. Co-morbidities obscure symptoms and changes to the DSM have, in many ways, increased the overwhelming misunderstandings of the disorder. In fact, many researchers ignore the DSM and focus purely on symptomatology

perpetuating the lack of interoperable information [234]. Current PTSD diagnostic instruments can be inaccurate and clinicians consistently disagree about the manifestations of the symptomatology while other well-respected scholars attribute a large portion of the symptoms as psychosomatic. For several years, controversial forensic psychiatrists have been reviewing military medical records and reverse diagnosing without patient input, which led to fines and loss of positions [235-236].

The studies conducted over the past decades do not form a cohesive body of evidence on effective treatment [32, 49]. Exploration of new therapeutic interventions have made this a popular disorder to promote funding and interest, however a measure of success has fallen short [7]. It merely takes skimming weekly PTSD headlines that consist of "groundbreaking" treatments that are ridiculous in theory, illegal in some cases, and others that can actually make posttraumatic symptoms worsen. Objective improvement measures are hampered without an established definition for recovery. It is not currently possible to differentiate trauma survivors who will recover naturally from those who will develop enduring symptoms [9]. Many of the failures have come from disagreements between definitions of terms and the inability to extract information from the clinical narrative of patients. A clear area of opportunity for improving the understanding of PTSD is the extraction of narrative text from its many sources in order to begin to measure treatment outcomes and variations in symptomatology.

The awareness of terminology as a rate-limiting resource outside the confines of medical science and informatics is generally low. Data from various formats and structures create heterogeneous information that is unusable if not interoperable. The diverse design structures enable ambiguity in terminology and foster misunderstandings when specified context is missing. Referencing standard terminologies often have limited coverage for identifying mentions of concepts within text [58]. Classification and representation in terminology coverage is often confusing, ambiguous, and imprecise [61]. The coverage of concepts necessary to accurately describe a domain can be overwhelming, costly, and difficult to achieve.

There is a well-defined need to put objects into classes or categories in order to describe the large amount of health and science data that must be managed. The conceptualized categories are representation of concepts used for human description. Shared understanding has been hampered because agreement on categorization of objects is not as of yet resolute [68]. Labels for these categories are necessary for communication in order to achieve the agreement. The increase in

user acceptance of terminology standards for each developed lexicon can dispel polysemy. The goals of each stakeholder relates to the ultimate ability to use and analyze their respective data. The terminology, medical, and scientific community must establish agreement via standards for trusted sharing of information [23]. There are gaps in the standard terminologies and the data elements required to document information. Opportunities are vast for improvement in sharing, and representing knowledge for analyzing highly specified domains in science and medicine [21, 62, 138-141].

Existing standard terminologies may have limited coverage and missing information with respect to concepts and their meaning in clinical narratives. There are opportunities to support expansion of biomedical knowledge by standardization and integration of terminologies into unified infrastructures [21]. Formal integration has been successful in efforts such as the UMLS but can also be achieved through interoperable structures sharing information and metadata. There are applicable structural problems requiring revisions in the conceptualization of these integration efforts [196, 198]. Every segment of the biomedical domain requires improved depths of knowledge coverage and consistent updates. Regardless of terminology application scope, ultimately comprehensive domain coverage will be required to support future biomedical, text mining, and artificial intelligence applications [77].

To include PTSD, there is an overall lack in mental health terminology knowledge bases with depth and breadth of coverage capable to support biomedical applications [14,131]. Existing vocabulary resources either contain high-level concepts and general biomedical terms or are made up of relevant concepts but with improper definitions or meaning for the sub-domain. A majority of the mental health domain is in need of a comprehensive terminology of concepts, their respected word senses, and lexical context according to usage. This vast vocabulary must include synonyms, antonyms, hypernyms, hyponyms, meronyms, etc., and keep up to date with the evolving and rapidly growing field of information. Mental health terminology standards need further enhancements to ensure interoperability between independent or discrete systems and compatibility of data for comparative statistical or analytical purposes. Standards also reduce lexical redundancies, ensures conformity assessment and usability [21, 61].

In mental health, two common standard vocabularies are found in the Diagnostic and Statistical Manual (DSM) and the International Classification of Disorders (ICD). The DSM is the standard for psychiatric diagnosis in the United States and the ICD is the international

standard diagnostic classification for epidemiological coding, health management procedures and clinical use for mental health [237]. The diagnostic status of PTSD by the DSM is controversial and still appears dissatisfying to many [238-239]. Expert identified issues inhibiting understanding include overlapping non-specific symptoms shared with other common mental health disorders, its broad definition of diagnosis, and how its pathology follows stress reactions to normal events. Even with recent updates to the DSM, the changes have not yet reduced confusion [240-241]. Focus on specific disorders individually but with interoperability considerations is a strategy to address this terminology gap. Putting efforts such as terminology aggregation, text mining, natural language processing (NLP), and data mining development techniques to the growing disorder of PTSD is a well-positioned opportunity to make an impact on improved understanding of the disorder [21, 23, 242].

Information is evolving and new research provides novel insights that require change in terminology conceptualization. There are numerous ad hoc design techniques that limit interoperability and capacity to reuse shared knowledge across specialized disciplines. There are opportunities in establishing novel development techniques. While the application for which a terminology was designed provides an accurate test, there are also opportunities to standardize evaluation in order to compare intrinsic characteristics and extrinsic value [21, 77].

As directly stated by D'Avolio et al., "very few clinical natural language processing (NLP) or information extraction systems currently contribute to medical science or care" despite decades of pragmatic implementations in diverse settings [243]. Performance of systems have shown promise in highly controlled settings but achieved success does not directly transfer to multiple settings with unforeseen variables. NLP and IE in the clinical and medical domain has lagged behind processing of other domains mainly because of limited access to shareable clinical data due to constraints that protect patient confidentiality. There are challenges in reduction of noise and false positives that can block the advancement of text and data mining. The future of both techniques is applications in personalized medicine development and translational research. This will require fostering of a community focused on shared data, annotation guidelines, annotations, and evaluation techniques [244]. The capacity of scientists to stay updated with research is threatened by the quantity of scientific literature being published [245]. There are opportunities for developing user interfaces, improving usability and interactivity, integrating tools and mining

resources [246]. Text and data mining have the potential to improve medicine and life science by generating conceptual insights currently not available [247].

Chapter 3

Vocabulary requirements assessment, knowledge acquisition, and development of a PTSD terminology framework

3.1 Introduction

The importance of standardizing healthcare terminology has long been established for providing structure to data for synthesis and sharing across the continuum of care [1, 2]. Utilizing terminology frameworks such as controlled vocabularies, taxonomies, and ontologies promote standards and improve communication by providing a formal representation of the entities in a specified domain [3, 4]. In addition, the utility of these terminologies in clinical or health information applications has not yet been fully demonstrated. In this chapter, we describe the development of a terminology framework in the medical sub-domain of post-traumatic stress disorder (PTSD). The American Psychiatric Association defines PTSD as a condition occurring from exposure to a trauma that impacts the physical integrity or life of the individual or of another person [5, 6]. It is considered normal for an individual to have a strong reaction to a traumatic event but the effects should decrease over time when the threat is no longer present. However, people with PTSD continue to experience extreme reactions and symptoms even after the trauma is no longer present [7]. According to the National Center for PTSD, 7-8% of the population in the U.S. will have a form of this disorder at some point in their lives [8, 9].

The prevalence of PTSD continues to grow, particularly in the military veteran population, as combat operations continue around the globe and as researchers begin to better understand the disorder and identify it in patients once previously missed [7, 10]. The current healthcare system is also not equipped to cope with this disorder. However, there is continuous increase of PTSD information and it is growing at a processing overload pace. Specifically, the challenge of the data stored in narrative free text contributes to the complexity of dealing with the disorder. From an informatics standpoint, there are significant opportunities to assist healthcare communities with improved PTSD understanding through initiatives of health information system development, information extraction (IE) and natural language processing (NLP) tools and

projects. Before these tools and applications are built, standards and interoperability and how to maximize collaboration must be examined [25, 151]. Consideration must be given to the various ways PTSD terms are defined and used in healthcare and the research community [21].

The availability of knowledge bases has been identified as one of the critical elements that can help to realize the benefits of information management and clinical decision support [11]. This is especially true for PTSD vocabulary that is subjective, ambiguous, and overlapping with other mental health disorders. There is a lack of collective terminology necessary to support the future of clinical applications, and informatics initiatives in the domain. Improvements to vocabulary must be made to address terminology issues such as data heterogeneity (various data formats and structures), disambiguation (multiple word meanings), missing and incomplete information. These improvements to develop a PTSD terminology framework requires a rigorous needs assessment to identify vocabulary and coverage requirements. The goal is to discover important user needs that can be translated into knowledge acquisition requirements. The requirements engineering process consists of specifications that define, describe, and unambiguously communicate the stakeholder needs [12]. This dissertation explores terminology development under the assumption that inadequate requirements engineering would inhibit quality of the overall project. The terminology development methods described in this paper are based on various stages of complexity involved within vocabulary synthesis. The aim of the methods is to provide the non-expert developer with a framework capable of supporting terminology application in multiple use cases and at varying levels of complexity. Appropriate documentation fosters transparency and replication of the development steps [13].

The requirements derived from the needs assessment are used to guide the development of the PTSD terminology framework. Once requirements are determined, a key step in the terminology development is the knowledge acquisition of terms and concepts [248]. Knowledge acquisition is the process of synthesizing a vocabulary by extracting, structuring and organizing knowledge from several sources [249]. This terminology framework is constructed in the form of a scalable controlled vocabulary incorporating a hierarchical classification structure with continuous additions of new concepts. The long-term goal is to apply structure to unstructured information in order to improve the understanding of PTSD by creating a collection of descriptive vocabulary. The coverage of the framework will be directed towards 1) variations in symptomatology, signs, and characteristics; and 2) therapeutic interventions. The aim is to make

assumptions of the disorder explicit and reduce the ambiguity in concepts that describe symptoms and treatments within the domain. The research objective is to populate PTSD symptom characteristics/clusters and treatment categories with concepts from existing terminology resources in addition to new concepts. The implementation of the terminology framework will support the development of text processing pipelines for information extraction in clinical narrative text.

3.2 PTSD Terminology Development

3.2.1 Post-traumatic Stress Disorder (PTSD)

Central to this research, a primary focus is on developing a PTSD terminology framework in order to identify, categorize, and structure knowledge within the domain. In this regard, this chapter provides a background for PTSD, introduces prior research concerning developing vocabularies in similar medical sub-domains, and presents relevant publications relating to requirements gathering for terminology structures. The American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders (DSM) defines Post-Traumatic Stress Disorder (PTSD) as a condition resulting from exposure to direct or indirect threat of death, serious injury or physical threat [5-6]. The events that can cause PTSD are called stressors and may include natural disasters, accidents, or deliberate man-made events/disasters, including war. Symptoms of PTSD can include recurrent thoughts of a traumatic event, reduced involvement in work or outside interests, emotional numbing, hyper-alertness, anxiety and irritability. It is considered normal behavior for an individual to have a strong reaction to traumatic event but the effects should decrease over time when the threat is no longer present. However, people with PTSD continue to experience extreme reactions even after the threat is no longer present [7].

3.2.2 Terminology Systems

While the prevalence of PTSD continues to grow, it is still largely under-detected due to the difficulties in diagnosis, volatility of symptoms and lack of effective screening in healthcare facilities. The insufficiency of adequate screening is an important public health initiative to be addressed [7, 10]. However, there are numerous programs and projects underway in various organizations and locations throughout the United Sates. The organizations have overlapping

terminology needs where if shareable, provide an opportunity for synergy [250]. There are also a large number of projects at local healthcare facilities that acquire, analyze, and archive clinical research data of the disorder [251]. A major gap inhibiting the research-to-cure time frame is that many of these initiatives are operating in silos instead of collaborating. These projects have many of the same goals, share the need for better access to high quality information, and must overcome interoperability issues. From an informatics standpoint, there are significant opportunities to assist healthcare communities with improved PTSD understanding through initiatives of health information system development, and through IE and NLP tools and projects.

Significant improvements have been made with managing health information and there is high efficiency with the handling of the vast collection of available peer-reviewed literature. However, the amount of knowledge has been growing rapidly and continues to become increasingly difficult to search [252]. Researchers must continuously find ways to improve the management of this information. Another area where the amount of knowledge and information is accumulating at a rapid pace is in electronic health records (EHRs). There is a growing scientific evidence base from the PTSD research and clinical communities, including genetic and functional brain chemistry biomarkers, predisposing demographic risk factors, psychological and functional assessment factors [8]. A large percentage of data is stored in narrative free text, especially for mental health patient records, and automation of this data is particularly challenging. Continued improvements to knowledge management must be pursued for supporting clinical information support systems to make use of medical literature and EHR documentation. Improvements could come in the form of automating clinical processes and developing tools to help clinicians and researchers better understand the disorder. Before building these tools, consideration must be given to supporting resources of terminology and the accuracy of concepts it contains. Without proper use of terminology, it can become a bottleneck for the deployment of innovation and the assessment of quality in PTSD research [20]. Opening that bottleneck will require understanding of elements of this text and leveraging it into an asset for healthcare and research [21]. A significant challenge to leveraging the information that can be developed from this data are its storage in many varying formats. A majority of the existing databases have overlapped data because they are built remotely and independently from one other [253]. All of this information is captured in different data formats that make it almost

impossible to understand existing relationships between complementary data. Many researchers need much of the same data but with different meaning or context providing an opportunity to improve understanding of PTSD with the initiatives of terminology development.

Publicly available knowledge bases are limited and those that exist are generally derived from large terminology structures that are too broad, limiting accurate term identification. Additionally, those implemented for biomedical applications are developed in silos to support specific local tasks, which are not shared and validated by subject matter experts in the research community [254]. There is a lack of a collective PTSD knowledge base with depth and breadth of coverage yet sufficiently detailed capable to support initiatives with text mining applications. Existing terminology sources either contain high-level concepts and general biomedical terms or are made up of relevant concepts but with improper definitions or meaning for the sub-domain of PTSD. The research community is in need of a comprehensive collection of PTSD concepts, their respective word senses, lexical context according to usage, synonyms, and acronyms [255].

3.2.3 Needs Assessment

Developing a terminology into a usable framework is expensive in both monetary, time, and data curator resources. A very general gap in available PTSD terminology was confirmed by a feasibility analysis as a part of this research. It also identified many potential users of a collective vocabulary structure and validated the benefits to this effort which were determined to be realistic and valuable. With limited resources for this project, the many potential users and the lack of existing PTSD terminology coverage require the implementation of a rigorous needs assessment for vocabulary synthesis. According to Kaufman et al., a needs assessment is a process to "identify gaps between current results and desired one, place the gaps in needs in priority order," then "select the most important ones" that can be addressed [256]. A benefit of undergoing a needs assessment that identifies stakeholders is that it assists in allowing them to support the project. This greater support increases the number of ideas generated as well as the diversity of ideas. It brings in perspectives of various professionals to gain a realistic status of how this research can best be approached to maximize the benefit of use and the appropriate focus to be effective [257]. Thus, vocabulary needs were gathered and prioritized as it was impossible for this research to satisfy all needs for all stakeholders identified. The newly

developed vocabulary should be judged based on whether or not the objectives and requirements derived from the assessments of these needs are met.

3.2.4 Requirements Engineering

Requirements that are overlooked or unclear at the start of a terminology development project can be costly and difficult to correct as the project progresses [258]. According to Nuseibeh and Easterbrook, requirements engineering is "a process in which stakeholder needs are acknowledged to find the answer for a specific problem" [259]. This process should be systematic description analysis translating the stakeholder needs into what the terminology framework should accomplish. Often the more mature domain software development provides analogous examples of requirements engineering that taxonomy, ontology, and terminology framework developers at the beginning stage can reference [13, 260].A thorough understanding of the processes that software developers utilize are critical to terminologists creating a novel vocabulary reference. Software engineering relies on requirements gathering as a formalized process of assessing user needs to inform tool design. The rigorous methodology typically gathers requirements through a full accounting of the needs of the end-user [261]. This requires the developer to spend considerable time researching disparate resources resulting in a collection of techniques that may or may not meet user needs. In order to overcome these obstacles, the traditional requirements engineering approach is supplemented with contemporary agile methods. Agile methodologies [262-263] are a family of popular software development processes that contains elements that are included in the development of the PTSD terminology system. Features of the methods listed in **Table 3.1** were applied to both PTSDO requirements engineering and terminology framework development.

Method
• eXtreme Programming (XP) • Scrum • Dynamic Systems Development Method (DSDM) • Adaptive Software Development (ASD) • Crystal

Table 3.1. Applicable agile methods for requirements development [264-269]

The aim is to implement requirements efficiently and address stakeholder needs through the application of the principles of iterative development methods [269-270]. These agile methods can be applied to the volatile needs of requirements engineering for terminology development. Changing requirements can be managed by developing incremental requirement specifications in a manner where stakeholders can provide frequent feedback as a greater understanding of terminology needs are understood [271].

3.2.5 Knowledge Acquisition

Knowledge acquisition is the process of extracting, structuring and organizing knowledge [272]. An important element in knowledge acquisition is the extraction of knowledge from relevant sources and its population of the terminology system. It involves concept identification from existing terminology resources, human domain experts, documents, or any other reliable source valuable to the user's acquisition goals. The acquisition of knowledge can be conducted throughout the development process and likely in perpetuity as new information is discovered and made available [144, 205, 273]. Terminology systems can support various data serialization formats such as Extensible Markup Language (XML), JSON, OGDL, YAML and CSV to list a few [274]. XML has been successful at fostering data interoperability [21]. A popular XML-based artificial language for which knowledge in a terminology system can be stored is the Web ontology language (OWL) [275]. According to the OWL Web Ontology Language Guide, it is "intended to provide a language that can be used to describe the classes and relations" between them that are inherent in documents and applications [276]. OWL is a standard for representing knowledge and requires a design environment such as the popular Protégé [277]. Protégé fosters the capture of entities and their relationships intuitively. According to Kola et al., Protégé OWL [278] is an extension that "allows knowledge modelers to capture knowledge in the OWL framework" [279]. It is the leading ontology editor across disciplines, with a community of more than 50,000 users, representing research and industrial projects in over 100 countries. Beside the support of OWL, Protégé includes support for exporting terminologies into a variety of formats such as RDF/S and XML Schema [21, 277].

The knowledge modeling methods to support acquisition can be classified into the three categories of manual, semiautomatic, and automatic. Manual methods elicit knowledge from the experts or other validated resources to input into the developed terminology. In automatic

methods**,** the roles of both the terminology developer and domain expert are minimized or eliminated. In reality, the majority of successful methods are semi-automated meaning the automation requires some manual input for accuracy and efficiency [144, 273]. NLP techniques [143, 196] can guide concept identification and maximize the accuracy of providing some degree of automation [280].

Rindflesch et al. [84] of the Semantic Knowledge Representation (SKR) group at the National Library of Medicine (NLM) utilizes techniques of knowledge acquisition. This group utilizes a methodology that supports the development of a visualization and automatic summarization tool referred to as Semantic MEDLINE [281]. The core technology for this application is the SemRep [196] NLP system developed at maintained at NLM. SemRep provides partial semantic interpretation of MEDLINE citations. This NLP system is rule based and supported by domain knowledge in the Unified Medical Language System (UMLS). The UMLS integrates multiple electronic clinical and biomedical ontologies and terminologies [282]. The techniques described by SKR group members Rosemblat et al. [198] enhance the current UMLS by augmenting the resource with domain-specific concepts and semantic types.

3.2.6 Concept and Term Overlap

There is a tremendous amount of ambiguity within PTSD terminology which makes acquiring and vetting by community experts critical for ensuring consistency across resource development. Collecting and standardizing PTSD data is challenging because of its dispersed language in published literature, clinical notes, and consumer health forums [283]. Terms are words used to refer to a concept. A single term may refer to many different concepts and synonyms occur when a concept has more than one term referring to itself. Because research areas in life science and healthcare often need more than one terminology to capture all needed context and semantics, the quantity of terms required grows quite large. These various data sources need to be unified in order to consistently share information. Concepts are available in many terminological resources and occur in many vocabularies with exact, similar, or contradictory meaning. The continued development of multiple terminologies propagates the data silo problems that are prevalent today [284, 285]. The goal of an interoperable terminology is to constrain it so as to converge with existing resources [286].

Reuse of existing resources promotes the development of terminologies that foster interoperability and are cost-effective in both time and money. Reuse can be determined by examining a term or concept's mapping or explicit reference. Overlap occurs when multiple terminologies contain but do not reuse the same term or concept [283-285]. This research distinguishes terms from concepts in order to investigate each respective overlap. Identifying the overlap is important because it portrays the significance of the lacking standards developers are implementing for future data sharing and exchange of information [286]. Kamdar et al. referred to this phenomenon as the overlap–reuse gap that developers must focus on minimizing [287].

3.3 Description of Aim 1

Aim 1. Assess PTSD terminology needs for the requirements, knowledge acquisition, and development of a terminology framework to support text mining tasks

3.3.1 Conceptual Model

Figure 3.1 describes the conceptual model of aim 1. Terminological needs are translated into requirements for ongoing terminology development and knowledge acquisition. The engineered requirements guide the development and is utilized as a reference for ongoing text mining pipeline implementation.

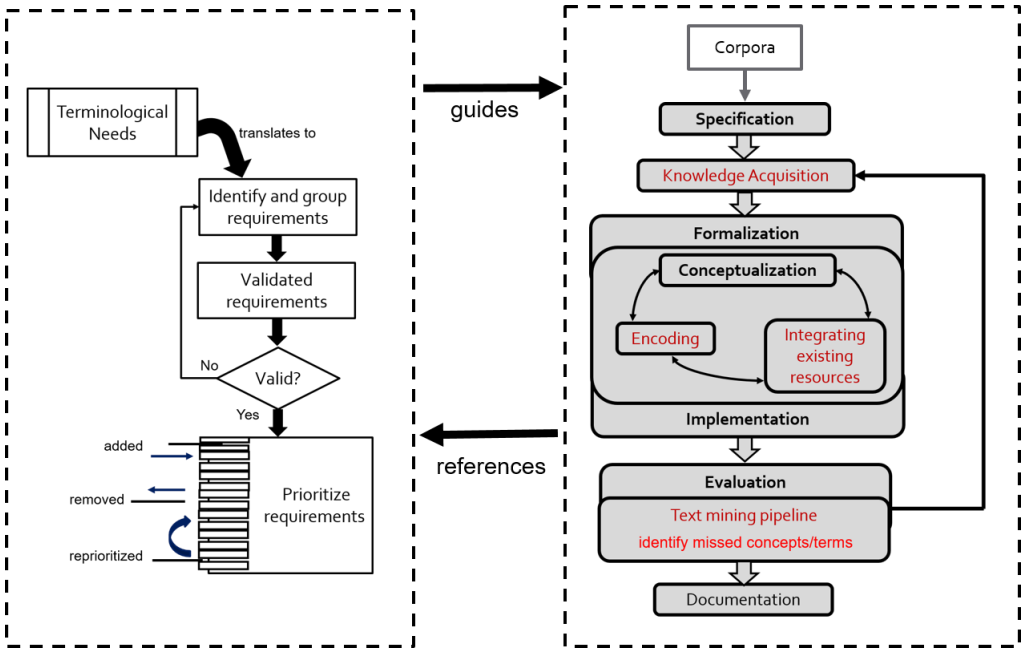


Figure 3.1. Conceptual model of aim 1

3.3.2 Research Questions and Hypotheses

RQ1. *Can sufficient PTSD terminology resources be identified in order to acquire salient concepts for expansion of domain coverage?*

Task 1: Perform manual mining of existing terminology resources

Task 2: Assess needs of stakeholders through iterative interviewing for requirements vetting and knowledge acquisition from experts

Task 3: Implement text mining pipeline for semi-automated mining of terms and concepts

Task 4: Examine knowledge acquisition saturation of concepts, terms and words from a range of terminology resources

RQ2. *Can the identification of PTSD entities described implicitly in unstructured data be automated in an iterative method to populate a structured hierarchical framework?*

Task 1: Examine concept overlap of acquired knowledge from available resources

Task 2: Examine term overlap of acquired knowledge from available resources

Task 3: Perform iterative domain expert validation of collected concepts and terms

Hypothesis 1 - Manual mining of existing terminology resources performed during needs assessment will produce a generic PTSD terminology system base to maximize stakeholder input

Hypothesis 2 - Manual knowledge acquisition of existing terminology resources will not produce the comprehensive needed concept coverage

Hypothesis 3 – An iterative semi-automated text mining approach to knowledge acquisition will determine when sufficient coverage is met in order to satisfy identified stakeholder needs

Hypothesis 4 – Concept overlap will exist by more than 50% in existing concepts identified from available resources

Hypothesis 5 - Term overlap will exist by more than 70% in existing concepts identified from available resources

3.4 Methods

Assembling a team is a critical step and deserves much attention. Successful terminology system development requires both PTSD expertise and in-depth knowledge of terminology framework capabilities. For this research, a great amount of attention to the solicitation of subject matter experts was undertaken in order to gather precise requirements to apply structure to unstructured text. The project aimed to research the capabilities and requirements of a vocabulary that will meet the PTSD caregiving, research, and biomedical application communities' information needs. Methodologies of needs and requirements gathering originate

from the fields of artificial intelligence, database development, and software engineering. Terminology development methods implemented are derived from the fields of controlled vocabulary, taxonomy, and ontology engineering.

3.4.1 Needs Assessment

The gaps of existing terminology resources with sufficient coverage to support PTSD text mining initiatives was confirmed in prior feasibility studies. A general terminology needs assessment is conducted in order to approach a more specific requirements gathering process with maximum focus. A needs assessment is a systematic process for determining gaps in current data and desired knowledge [288]. Kizlik states that the "need can be a desire to improve current performance" or address deficiencies in available resources [289]. The focus, shown in **Figure 3.2**, is led by accurate identification of stakeholders since not including the correct support personnel could derail the project. The task is to identify key personnel that can support the management of a needs assessment and a more detailed requirements gathering. It is important to find individuals that are committed to engagement and support follow-up questions implemented in the iterative needs assessment methodology. Timely reports to top management and other important stakeholders, with opportunities for interaction on major issues, are also critical [290].

The stakeholder interviews are focused to identify needs that are currently unsatisfied and the features that obtainable vocabulary structures do not fulfill. This is meant to assist terminology development in the decision to create vocabulary from scratch or enhance features of existing resources [291]. The desired outcome is to reach a consensus on the needs of greatest importance to the target group. Interviews are conducted in a manner to assist in the brainstorming of major concerns for each need and their respective verification. The interviews of identified stakeholders determine the kinds of information to clearly define the need and where to get the data. During these sessions, stakeholders were also introduced to an online spreadsheet of ongoing query collection associated with their identified needs. Lastly, the needs are prioritized according to feasibility and stakeholder consensus [288].

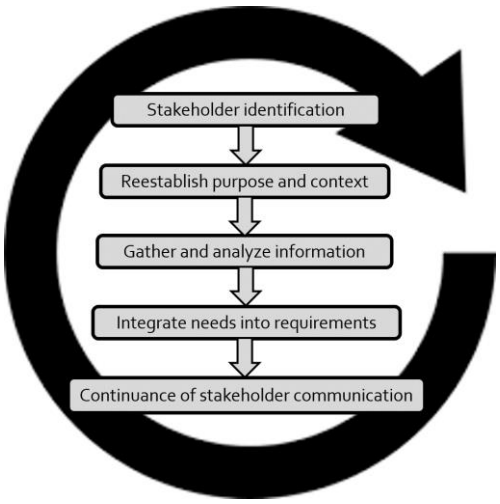


Figure 3.2. PTSD terminology needs assessment methodology

3.4.1.1 Stakeholder Identification

De Vries et al. [292] provides a method for stakeholder identification and classification in order to maximize participation. The method consists of a set of search heuristics to identify all relevant stakeholders, and a typology that can be used to differentiate between essential and less important stakeholders. Based upon stakeholder theory [293], participants are contacted and informed of the project in order to prepare for the needs assessment. De Vries et al. states that "stakeholders should be identified that can affect and are affected by" development of the terminology. The authors classify search directions for locating potential stakeholders of which end-users, researchers, faculty members, and members of PTSD representative organizations are relevant for this research [292]. Within each stakeholder group, all specific roles available were identified for contribution discovery.

For this project, a stakeholder is defined as any expert with informed input regardless of their seniority [294]. Semi-formal interviews were conducted with identified stakeholder experts to discuss the implementation of medical terminologies for text mining. For maximizing productivity, four questions were posed in order to prioritize interviews:

1. Does the stakeholder have a thorough understanding of the project space which can evaluate the needs that a vocabulary and terminology structure can address?

2. Does the stakeholder have a fundamental impact on their organization that can affect change?

3. Can the stakeholder identify what other users in their profession value as necessary to improve outcomes?

4. Does the stakeholder have a thorough understanding of the workflow of their organization to identify relationships and other potential stakeholders [295-296]?

3.4.1.2 Stakeholder Interviews

Semi-formal interviews were implemented in order to assess needs. The main objective of the interviews was to consolidate the wants and needs of the potential users and to answer the primary question of: "how should vocabulary be structured in order to meet the needs of its users?" Stakeholder concentration was focused toward determining capabilities that PTSD terminology must support. At a system-level, brainstorming was encouraged to identify what applications that implement the vocabulary should be able to accomplish. A sample of key questions that participants were guided to answer are:

1. What is the purpose of a collective PTSD terminology?

2. What is the underlying goal of a successfully developed PTSD terminology structure?

3. How can stakeholders identify when a PTSD terminology need is met?

These questions assisted participants in expressing a need properly for the development of requirements. As in similar needs assessment undertakings, the strategy is to focus on the problem of PTSD vocabulary not existing rather than solutions provided by its creation. The tasks described in these needs assessment methods were to determine the feasibility of a PTSD terminology structure development and as preliminary work in the planned requirements engineering processes. Its implementation is to understand why a particular solution may be required and collect the constraints of subsequent requirements engineering [291].

A SharePoint spreadsheet was maintained in order to collect questions stakeholders desired to be answered for research, patient care, or data processing. The collection of queries was necessary for the terminology system and text mining pipeline to answer in order to address respective needs. The spreadsheet fostered continuous user involvement and served as a basis for integrating needs into defined requirements for terminology development guidance.

3.4.2 Requirements Engineering

Requirements engineering includes all activities related to elicitation, discovery, analysis, specification, validation, documentation, and management [13]. It can also involve some levels of modeling, ranging from the creation of use case models to more detailed collaboration with system architects and designers. No explicit methodology exists for requirements engineering for the development of a controlled vocabulary, terminology framework, or ontology. However, techniques from software development can be adapted from several methodologies in order to meet the objectives of this research [297-298]. The needs assessment was used as a terminology framework pre-development feasibility study to facilitate the requirement gathering sessions, technology integration, and compatibility analysis. It was determined that a successful requirements engineering undertaking must be highly flexible and concentrated in order to maximize success [13, 260]. Thus, the requirements gathering phase was focused on answering the question, "What do the users need from a PTSD collection of terminology for text mining?"

Similar to Lopez et al. that utilized software engineering techniques in ontology development [299], this dissertation implemented a comparable approach for collecting terminology requirements. As shown in **Figure 3.3**, a contemporary agile approach for gathering, using and evaluating requirements was utilized in order to take advantage of implementable pieces of existing established methodologies in both ontology and software development [13. 138-142]. Requirements discovery is undertaken to define/refine the terminology purpose and scope, identify all users, and to elicit their needs. Throughout this process, stakeholder needs and wants were separated and prioritized. The agile process allows for prototype terminology systems to be tested against a selected corpus during development. This design approach was governed by a requirements management approach that included requirement development traceability and their respected changes [139-140, 300].

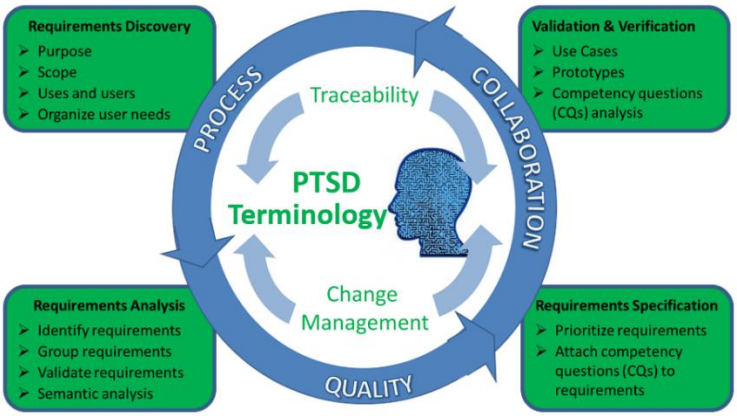


Figure 3.3. PTSD terminology requirements engineering methods

Because there was no clear picture of what the terminology system should contain, agile methodology techniques were utilized in addition to traditional development methods. The intention of the agile approach to requirements engineering is to develop the terminology, requirements, and implement text mining pipelines iteratively [269].

3.4.2.1 Requirements Elicitation

Requirements elicitation is used to gather relevant requirements for the development of the terminology framework. User needs were discovered for the PTSD terminology development by exploring focus groups, brainstorming, non-structured interviews, shared idea web document, use case/scenario-based analysis, and prototyping. These methods also facilitated both concept identification and the requirements engineering process. To guide requirement gathering sessions, contributors were given a handout providing guidance of focus toward 'SMART' requirements [301]. **Table 3.2** provides a summary of SMART including the attributes and their respective definition.

Attribute	Definition
• Specific	formulate requirements as specific as possible
• Measurable	requirements need to be measurable
• Actionable	requirements must be executable
• Realistic	requirements must be realistic and feasible
• Time-bound	when requirements are fulfilled must be clear

Table 3.2. Attributes and meaning of SMART requirements methods [302]

During the brainstorming session and interviews, stakeholders were asked to consider input on a comprehensive and inclusive level. The interview questions are attached in **Appendix A**. A sample of the semi-formal interview questions include:

1. What are common strategic goals in your PTSD research and what types of data do you typically need in order to achieve these goals?
2. What are the kinds of analysis you perform or wish to perform on the data you obtain?
3. What do you not like about the terminology resources or knowledge bases you use/what are their respective shortcomings?

Stakeholders were directed to not only pay attention to the functional requirements, but to the non-functional requirements as well. The methods ensure the goals and objectives of the framework remain correctly understood by all stakeholders [13, 269]. At the end of sessions, all the requirements gathered were fitted within the scope of the project for examining feasibility [270]. At this stage, competency questions began to be collected by asking each of the stakeholders to identify a set of queries relevant to address stakeholder needs. Common research identified includes both qualitative and quantitative analyses supported by terminology standards. Research goals identified include: 1) aligning terminology with the recording and reporting a patient's care at varying levels of detail; 2) ease clinical proceedings and adopt current best practices; and 3) describe unambiguously the care and treatment of patients. Data types used in research are acquired from experiments/clinical trials, recording well-defined events, obtaining relevant data from management information systems, electronic health records, administrative data, claims data, disease registries, health surveys, and clinical trials data. Ad hoc modifications to various terminology structures prevents users from cross searching multiple repositories, cross-sectoral resources and interdisciplinary material. From the list, it can be determined whether the terminology resource is able to answer these queries in enough satisfactory level of detail. This list of questions assists in developing the scope of the terminology framework [134]. From the list, it can be determined whether the terminology resource is able to answer these queries in enough satisfactory level of detail. This list of questions assists in developing the scope of the terminology framework [134, 303].

3.4.2.2 Requirements Analysis

Requirements analysis gathers those identified in the elicitation phase and groups them into a coherent structure for review. Several processes are implemented in order to ensure the quality of the gathered requirements. The quality criteria used for analysis in the PTSD terminology system are shown in **Table 3.3** [258]. Each requirement must accurately describe the required functionality, and must be technically correct. Completeness necessitates that each statement must fully describe the functionality to be delivered. The description must be sufficient for the developer to understand and implement it. Consistency means each statement should not conflict with its source requirement at a higher-level. The traceability of requirements is ensured by using standardized templates and rules for the documentation of requirements. Collaborative support from stakeholders ensure the relevance of the conceived requirements. Verification along with validation is conducted in the preceding steps which tests accuracy through implementation. The requirements that are needed continue to this next step while those that are not useful are discarded. The requirements analysis involves many stakeholders throughout development of the terminology in order to garner more insight into vague requirements.

Quality	Criteria
Completeness	no gaps in textual content and functionality fully understood
Consistency	no conflict with existing requirement
Traceability	source objective easily identifiable
Relevancy	necessary and benefit to terminology framework confirmed
Unambiguity	no question in interpretation/comprehension
Verifiability	tested to be true

Table 3.3. Quality criteria for requirements analysis [258]

3.4.2.3 Requirements Validation

Validation is the process of confirming the completeness and correctness of requirements which is verified testing. Techniques include prototyping, test case generation, and formal/informal reviews. It is important that this step in the process contain a conformation loop where domain and ontology experts provide feedback that the requirement is correct and this is documented for future terminology users. Use cases can validate whether this requirement has been correctly implemented. Reviews should be held regularly while the requirements definition is being formulated. Each must add some value to the terminology system as determined by the

stakeholder needs. Prototypes can provide clarification to stakeholders to identify problems with the requirements. As a prototype version of the PTSD terminology system is made available, testing of stakeholder competency questions can be conducted with implemented text mining pipelines. Competency questions will establish concepts needed for coverage as well as necessary relationships needed between identified concepts and serve as the litmus test once the system development is complete. The goal of this task is to identify possible conflicts between questions and contradictions in the identified queries. Users and domain experts carry out this task taking as input the set of grouped questions for deciding whether valid or not.

3.4.2.4 Requirements Specification Documentation

A requirements specification is a comprehensive description of the intended purpose and background for the terminology system under development. A summary of the requirements is attached in **Appendix C**. In this step of requirements engineering, the requirements are organized into a complete document to describe the functional and non-functional capabilities. It also defines how the terminology system should interact with various text mining applications. Secondly, requirements are prioritized in order to balance the project scope against the constraints of time and quality goals. Lastly, competency questions are attached to requirements and used for planning the terminology system development. The goals of the document include a means to describe the scope, facilitate reviews, and provide a platform for ongoing refinement via specifications enhancements [139-140, 142].

Displayed in the agile requirement and prioritization development **Figure 3.4**, stakeholder needs are translated into terminology requirements. Each requirement is analyzed and validated according to the methods previously described. Next. the stakeholders have the ongoing tasks of updating the priority of requirements to the scale of: 1) essential; 2) conditional; and 3) optional. Stakeholders were asked to assign a label to each requirement which were reviewed in a group brainstorming session in order to confirm the consensus of the label. The status of requirements change as needs evolve and new requirements are added according to the scrum management technique [265]. The strategy is to drop or defer low priority requirements to a later release when new, higher priority requirements are accepted or other project conditions change. Throughout the process, the requirements are prioritized based upon recommendations of the stakeholders and decision-making authority of the recognized project champion. When attaching competency

questions to requirements, the task output is to place each competency question into a group and attach each group to a set of requirements [264-265].

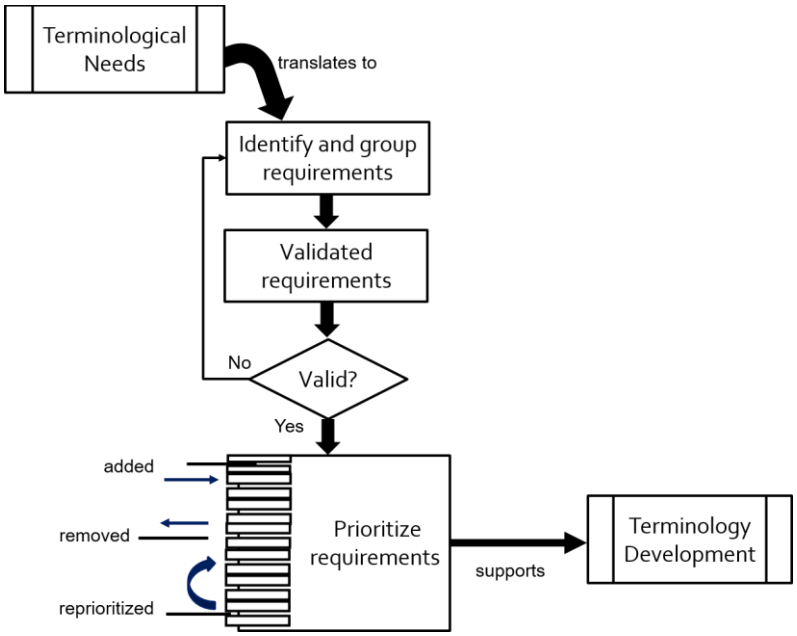


Figure 3.4. Scrum (agile) iterative requirements management [139, 264]

With the added iterative process as shown in **Figure 3.4**, the agile development applied to requirements is highly flexible [139, 264]. According to these Scrum methods, the requirements at the top of the stack have higher priority thus modelled in greater detail than the lower priority requirements at the bottom of the stack. Items can be added, reprioritized, or removed at any time. Each iteration will implement the requirement with the highest priority after each implementation iteration. Stakeholders have the responsibility for making decisions about changing requirement items in a timely manner [264].

3.4.2.5 Requirements Management

Once requirements are put in writing they become subject to management activities. These techniques include identification, traceability description, change, and version management [12, 13]. Each requirement should have unique identification in order to assure that each can be traced to other requirements, terminology information, as well as identifying and capturing the source of each requirement. Requirements change management adopts the agile principles of

accepting change input, addressing it, then continuing project development [269]. Formatting of the document itself maintains a version annotated by document dates, version numbers, revision, and substantive changes [140, 262-263]. A cumulative list of changes is kept to identify the changes from the preceding document versions.

With some regards, requirements engineering also happens throughout the entire project. The identification of important features of each requirement is the benefit of applying agile techniques to traditional requirements engineering [262-266]. Typically, terminology developers will begin by acquiring concepts in the domain of focus which contrasts with software engineers that begin with a requirement specification of the application. This research proposes a methodology that balances these conflicting philosophies. The goal is to achieve progress while maintaining a level of formality. Throughput the project, the most important requirements are identified according to priority [269]. The method is intended to assist in the management of the often competing and sometimes conflicting needs of the stakeholders. Locating a champion that the stakeholders respect can foster user involvement, assist in the decision making, and prioritize the requirements. The champion position requires high involvement and a thorough understanding of the PTSD terminology project for change management decisions [269].

3.4.3 Terminology System Development

The PTSD terminology is herein referred to as PTSDO, its designated PURL (Persistent URL) associated with http://purl.bioontology.org/ontology/PTSDO. The PURL server and naming classification is to enable identification and sharing of persistent URLs for ontologies, concepts, and mappings, etc. The acronym PTSDO was chosen as future implementation will be to convert the PTSD terminology structure into an ontology for supporting semantic web applications.

PTSDO draws upon a hybrid of design approaches to incorporate the disparate knowledge bases and terminology resources. Initial design will be based upon the Noy and McGuinness Ontology Development 101 [138] that provides an approach to create a generic base of the knowledge representation needed for the design of the domain model. A generic base allows the design to be more inclusive of available reusable resources without making assumptions that hurt the design. The building process consists of iterative steps, namely determine the domain and

scope, consider reusing existing terminology resources, enumerate important terms, then define the class hierarchy or “is-a” relations [75-76, 138-141].

The primary methodology for the PTSD terminology system will be the Methontology [142] methodology developed by Fernández et al. in the Laboratory of Artificial Intelligence at the Polytechnic University of Madrid. It has roots in the main activities identified by the software development process of IEEE 1074-1995 and knowledge engineering methodologies. The Methontology framework enables the construction of terminologies from scratch, reusing other terminology resources as they are, or by a process of reengineering them. As shown in **Figure 3.5**, it includes a life cycle based on evolving prototypes and particular techniques to carry out each activity. The process identifies which tasks should be performed and includes stages developed by Pinto and Martin [76] to include specification, conceptualization, formalization, and implementation of the terminology system. The life cycle identifies the stages through which the terminology system passes during its lifetime, as well as the interdependencies with the life cycle of other knowledge resources [75-76, 138, 142].

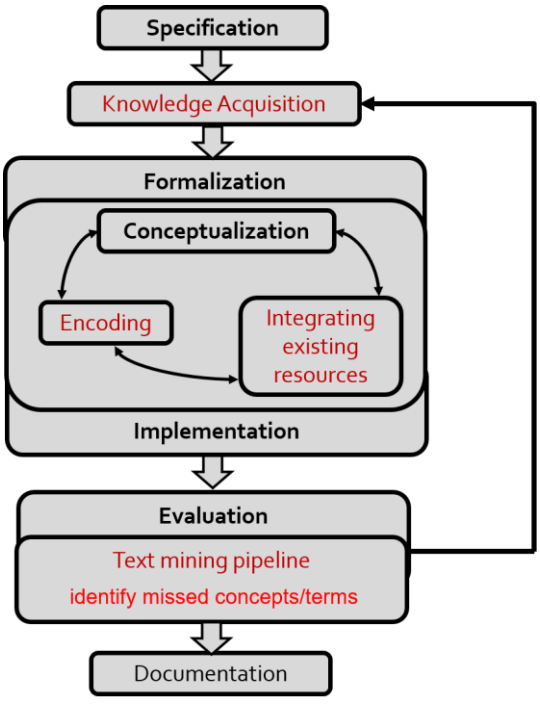


Figure 3.5. Multi-stage engineering development lifecycle [142]

Specification is the stage that determines the scope and the goal of developing the PTSD terminology system [76]. The stakeholders are determined from the requirements engineering initiatives but their information needs are expanded upon in order to refine the system's scope. During this stage, objectives are identified and a communication plan is established in order to keep stakeholders informed. Upon completion of the specification stage, a thorough understanding of why the terminology system is being built and each of its end-users is established. The collection and organization of the relevant domain concepts to be included is a part of the conceptualization stage [76]. The primary steps in this stage are vocabulary collection from identified resources and data analysis for validation of identified concepts to include their external references. If identified concepts do not include adequate definitions, their production will be required which must include precise unambiguous textual descriptions. It is important to note that the specification and conceptualization stage is primarily influenced from the requirements gathering tasks.

Formalization establishes hierarchical (IS_A) relationships in order to build the terminology system for use in the implementation phase [76, 198]. Coding techniques are implemented in order to in order to ensure explicit representation of all knowledge acquired. Concepts derived from the conceptualization stage must be integrated into existing resources via some mechanism of external reference. Symptom concepts are arranged according to clusters defined within the Diagnostic and Statistical Manual of Mental Disorders (DSM) [5,6]. The VA/DoD clinical practice guideline for post-traumatic stress management is consulted as a treatment concept arrangement reference [34]. The established classes for modelling the PTSDO are shown in **Figure 3.6a** and **Figure 3.6b**. Further evaluation will lead to the addition of new concepts and the contextualization or redefinition of existing ones for appropriateness in the PTSD domain.

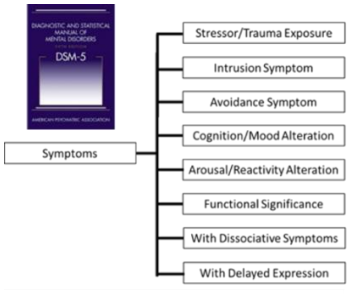


Figure 3.6a. Diagnostic and Statistical Manual of Mental Disorders, 5th ed.

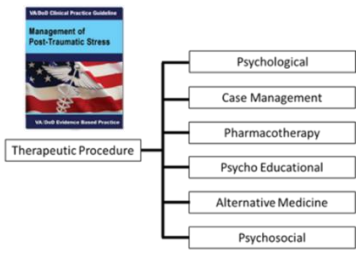


Figure 3.6b. VA/DoD PTSD clinical guidelines

The PTSDO is built using Protégé 4.8 [277] and represented in the Web Ontology Language (OWL) [275]. A thorough description of concept-mining is described in the knowledge acquisition section below. Formalization is the formal representation of knowledge. The PTSDO utilized Basic Formal Ontology (BFO) [129] as the upper ontology in preparation for future expansion towards the PTSD ontology. Coverage analysis by domain experts validates the intrinsic completeness and the correctness of the terminology system [142]. Implementation is known as converting the formalized knowledge into a machine-processable language [76, 198]. PTSDO must be compatible to support the implementation of multiple text mining pipelines. Evaluation, to include maintenance, is performed using a semi-automated technique which is described in **section 3.4.4** [138-142]. Similar to the engineering of requirements, the terminology system also applied agile methodology techniques to its development. Beck et al. describes agile software development by twelve principles of which the six listed in **Table 3.4** are relevant and adaptable for the terminology development involved in this research [264].

Principles
1. Continuous delivery
2. Adaptability to change
3. Accessibility to terminology
4. Champion for stakeholder motivation
5. Quality design
6. Prioritization of knowledge acquisition

Table 3.4. Development standards for PTSD terminology system [264]

The first standard applied from Beck et al.'s principles in **Table 3.4** is continuous and early delivery of the terminology. Following the traditional ontology development approach, gathered PTSD concepts from existing resources are aggregated immediately at the outset of the project. The required tasks for gathering of this knowledge acquisition discussed in **section 4.4** is completed in concurrence with requirements gathering. The second principle applied is the welcoming of altering requirements, even late in the creation process. Third, access to the latest version of the terminology was available to stakeholders throughout development and changes were made available immediately in order to gather feedback for incorporation into the project.

Next, the fourth standard is identifying a champion for research cohesiveness. The champion motivated the stakeholders for frequent feedback from those involved in the project. The saturation of acquired concepts discussed in **section 4.4** were the primary measure of progress. The fifth standard developed from Beck et al.'s principles is incessant consideration to quality design achievable through technical adherence to details of the terminology. Lastly, a standard of prioritization was maintained in order to identify the most salient terminology to be inputted into the system. Priorities maximized the amount of work completed and kept focus on coverage essential to requirements identified [264]. As stated by these standards, agile system development focuses on frequent adaptable delivery, close collaboration between stakeholders, and coping with changing requirements for quality and prioritized design. Terminology system creation can either be manual, automated, or a combination of both. It involves analyzing context, content, and users within the defined scope. Standards and guidelines help ensure classification consistency, an important attribute of a quality content management system engineering process [266-271].

3.4.4 Knowledge Acquisition

To develop a terminology framework that can be used to support NLP tools, the vocabulary used by domain experts and researchers must be obtained and defined. This vocabulary must be focused in order to provide a framework for a foundation upon which to build NLP-based concept extraction. Luther et al. states that a base structure of the terminology framework is developed from knowledge acquisition of relevant terminology. It is then followed by identification of synonym terms, formation of concepts, and hierarchical organization of concepts [255]. The terminology system is populated with content by a combination of manual and automated techniques from the vocabulary resources described in **Table 3.5**. The concepts and hierarchical relationships are derived from clinical guidelines, medical literature, controlled vocabularies, focus groups, cognitive interviews with providers, and annotation of clinical notes. These vocabulary resources are analyzed for high value candidate terms to add to the terminology collection.

For each candidate concept and term, existing terminology resources are referenced in order to apply reuse. Concepts are first mined from those accepted and submitted for acceptance to the OBO Foundry [128]. Next, the UMLS [24, 282] is searched followed by the NCBO Annotator

[195]. Identified concepts are referenced via an Internationalized Resource Identifier (IRI), explicit references (*xref*), or a Concept Unique Identifier (CUI) mapping.

- Diagnostic and Statistical Manual of Mental Disorders
- VA/DoD PTSD evidence-based practice management guidelines
- SNOMED-CT terms
- PILOTS (Published International Literature on Traumatic Stress) thesaurus
- PubMed literature
- Veterans healthcare administration corporate data warehouse
  - Adjudicated terms and spans of text
- PTSD reference documents
  - VA/DoD Evidence-Based Practice Working Group findings
  - Training materials from VA TBI-PTSD course
  - Wilson's *The International Handbook of Traumatic Stress Syndromes*
  - Colodzin's *Trauma and Survival: A Self Help Learning Guide*
- Specialty journals
  - *The Journal of Traumatic Stress*
  - *PTSD Research Quarterly*
- Books
  - Kardiner's *Traumatic Neuroses of War* and *War Stress and Neurotic Illness*
  - Kleber's *Coping With Trauma: Theory, Prevention and Treatment*
  - Peterson's *Post-Traumatic Stress Disorder: A Clinician's Guide*
- Combat action and casualty reports

Table 3.5. Terminology resources for system population

For semi-automated concept and term mining of these resources, methodology from the National Library of Medicine's (NLM) Semantic Knowledge Representation (SKR) group's Semantic MEDLINE development is applied [198]. The methods described here are semi-automated linguistic techniques that identify relevant terms from text based on corpus-driven concepts. While SKR implements NLM's SemRep [196] NLP system for semantic predication identification, these same methods are implemented with cTAKES [200] NLP system for concept identification. This method implements an iterative technique to support the evaluation and maintenance stage displayed in **Figure 3.5**. An outline of the semi-automated knowledge acquisition steps is shown in **Figure 3.7** for identifying core PTSD information to include applicable concepts, terms, and synonyms.

The gathering of a corpus input shown in **Figure 3.7** is the first step for the terminology system concept discovery process. Of the corpora collected, 10% is set aside for testing coverage of the developed terminology. The textual input is processed with the text mining pipeline supported with PTSDO converted to a dictionary reference. The processing represents the automated technique as described by the SKR methods [198]. Next, manual linguistic analysis is performed on the processed text in order to determine whether relevant PTSD terms were identified correctly. If all concepts are identified with cTAKES supported by PTSDO, the next textual input is processed iteratively. When a term is not identified, the PTSDO is researched in order to determine whether the string of text is represented within the structure. If it exists within PTSDO, NLP modules within cTAKES are adjusted in order to reevaluate term identification. Otherwise when the identified term is not in the dictionary, existing terminologies and ontologies are referenced as previously described in order to determine availability in an existing resource. Identified existing terms are incorporated into the correct hierarchical representation of PTSDO utilizing its external reference. Else, the identified PTSD knowledge string is created de novo and incorporated into the terminology system. After preferred terms, and variants are collected, their hierarchical relationships are established. For each term, analysis of whether a broader or more general term exists is researched. Each entity is also analyzed to determine whether more specific terms exist or are required to arrange within the terminology structure.

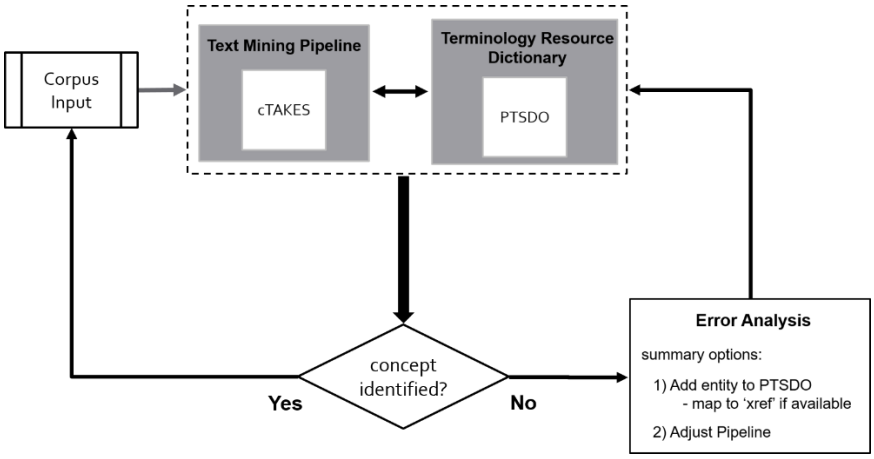


Figure 3.7. Knowledge acquisition for PTSD terminology [198]

The generic base of PTSDO is converted to a dictionary in order to support named entity recognition in the text mining pipeline cTAKES [200]. cTAKES is an open source clinical NLP

platform built with the Unstructured Information Management Architecture (UIMA) engineering framework [169, 180]. Components of the system include modules for preprocessing, sentence detection, tokenization, lemmatization, part-of-speech tagging, shallow parsing, dependency parsing, and named entity recognition (NER) [200]. By configuring PTSDO into a compatible structure, a dictionary-based approach is implemented in order to identify signs/symptoms and therapeutic intervention entities for acquiring knowledge. Concepts and terms not identified with the system due to not being included within PTSDO is recognized as a candidate entity. The identified electronic corpus available for knowledge acquisition displayed in **Table 3.5** is processed through cTAKES which outputs found terms. A regular expression module is added to the framework, shown in **Appendix F**, to increase entity identification through stemming. Knowledge acquisition of candidate terms are identified via an extraction script, however, the visual debugger of de-identified processed data is shown in **Figure 3.8**. Shown is the visual automated annotation output which is linked to XML output utilized for candidate term knowledge acquisition.

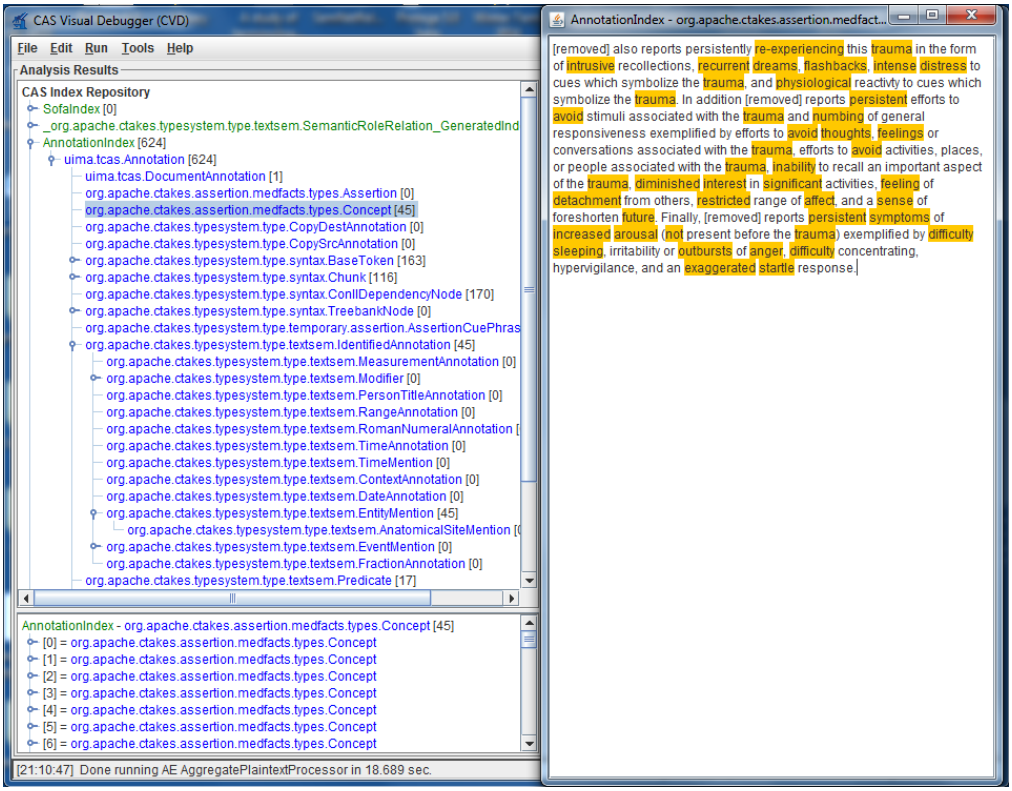


Figure 3.8. cTAKES visual debugger of automated annotation

3.4.5 Concept and Term Overlap

The techniques to examine concept and term overlap are applied from Kamdar et al. in order to define when overlap occurs [287]. A term is the word or phrase that is used for an object or idea. A concept can be a term itself but is also the idea itself or meaning of the word. A term usually has a preferred label, other labels, synonyms, and other properties. Each concept or term is defined in a source terminology and can be imported into other resources for fostering interoperability among resources. A concept or term is considered reused if it is available in two or more accessible resources and linked via an Internationalized Resource Identifier (IRI), explicit references (*xref*), or a Concept Unique Identifier (CUI) mapping. **Figure 3.9** shows examples of the various means of reusing a concept or term utilized in PTSDO. The types displayed include examples of reuse via an IRI, xref, and CUI [287]. The concept of anxiety in the national cancer institute thesaurus and SNOMED-CT are mapped to the same concept unique identifier (CUI). Aggressive behavior is defined in the gene ontology, reused in the neuro behavior ontology using the same IRI, then imported by xref into the emotion ontology.

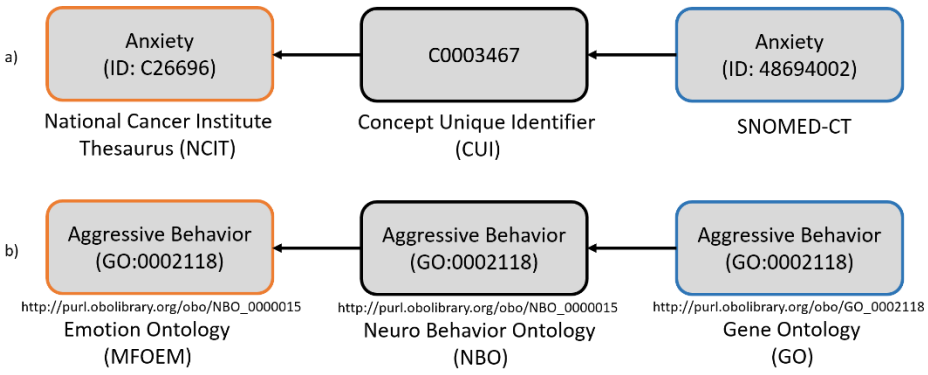


Figure 3.9. Example of concept or term reuse via IRI, xref, and CUI

A concept or term overlap is considered to have occurred when similar entities identified by either their preferred label or synonym are located in two or more terminology resources. Overlap is identified by terms or concepts created de novo in multiple sources but do not share a linkage or mapping. In order to calculate concept/term overlap, an entity-terminology matrix was maintained for each concept/term acquired and retrievable from an existing resource. A concept/term was considered reused when it appeared in at least two terminology resources identified by a matching external reference [287].

3.5 Results

3.5.1 Needs Assessment

The users along with their needs were grouped in order to better impact the objectives that each set of needs sought to address. The various categories of stakeholders generally evaluated their needs with the same set of criteria. As shown in **Figure 3.10**, end-users and stakeholders were categorized into clinicians, researchers, healthcare administrators, and a final category of experts in terminology and applications development.

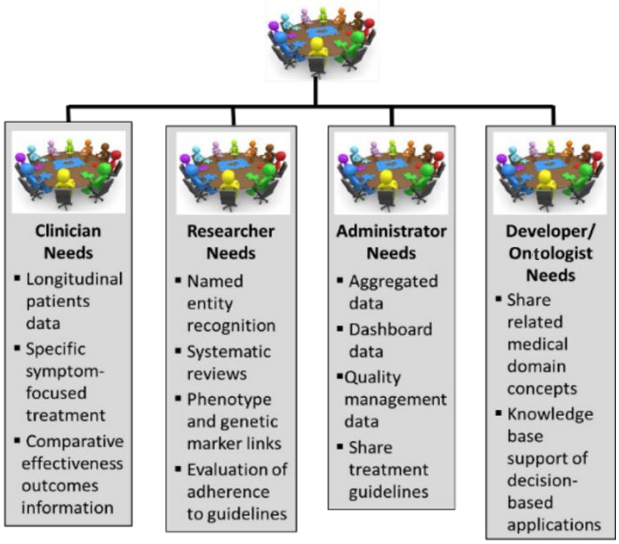


Figure 3.10. Identification of end-user and stakeholder needs

Clinicians identified a need to quickly find longitudinal information of their patients. Longitudinal information includes problems, allergies, medications, diagnoses, recent procedures, and laboratory tests that contribute to treatment and understanding of PTSD care. This group of users also identified needs for improved understanding of symptomatology and finding focused treatments to specific symptoms that traditional guidelines are failing to improve. Clinicians require more efficient methods of identifying evidence against using certain medications found to be contradictory for use with other medications in order to inform patients.

Another category of users and stakeholders is researchers to include scientists and investigators working with textual features of medical information. A need identified by essentially every user is the capability to aggregate PTSD treatment and symptoms concepts from examined corpora. There is also a desire for mechanisms to review treatments in order to compare effectiveness and identify therapeutic gaps in knowledge. Another need is locating

evidence of proven links between any documented phenotype and genetic markers related to PTSD. Lastly, there is a prevalent theme related to increasing the research on evaluation of adherence to guidelines for recommended practice.

Health care organization administrators have a need for implementing automated clinical documentation processing and support for reporting of aggregated patient status. This is precipitated by the necessity of accumulated treatment data for quality management. Regional and national administrators have need for policy decision support and efficient instruments to share treatment information between facilities. Across the board, professionals desire aggregated patient information for updating reports and dashboards to keep decision makers informed. They seek after mechanisms for exchanging data between districts, regional, and national facilities. An example commonly given was the exchange between researchers interested in investigating findings of new and innovative therapeutic strategies. Administrators are concerned with improving efficiency, better understanding patient behavior, and other health care economic factors. Application and software system developers stated, "it would be tremendously valuable for a dedicated PTSD knowledge base that could be implemented for supporting decision-based applications." Terminologists depicted a priority of sharing linked information. PTSD-domain experts desire resource collaboration and knowledge sharing. In the process of identifying stakeholder needs, many of the memberships among these groups overlapped.

The success and failures of the processes utilized in the needs assessment influenced the approach taken with the stakeholders input for the remaining work. The feasibility findings determined that requirements gathering via traditional software development life cycle methods would not be flexible enough. It also became apparent that agile development methods would not be focused enough. Based upon this assessment, a hybrid approach that applied appropriate techniques of each methodology types to requirements engineering and terminology system development is undertaken for this project. Competency questions gathered from stakeholders are primarily focused on entity identification of symptom and treatment concepts. A selected subset utilized in this project are attached in **Appendix B**, and include concept recognition questions such as:

1. Is this a symptom concept?
2. Is this a treatment concept?

It was determined that no text-mining system currently being implemented is able to identify all mentions of these concepts within clinical or biomedical text. Other questions identified include:

1. What are the relations between symptom concepts and treatment concepts?
2. What is the collection of symptoms that co-occur together?
3. What is the most effective treatment for a specified symptom?
4. What are the DSM categories for the collection of symptoms that co-occur together?
5. Which patients have a depressive episode predated by nightmares?
6. What are the symptoms of a PTSD patients with comorbid TBI?

These competency questions assist in the semi-formal modelling of the PTSD terminology system and support the necessary types of concepts to be included as well as ascertain their contextual description. The findings determine a priority of named entity recognition for aggregation of symptom and treatment concepts. This concluding need for concept recognition is also a prerequisite for answering the more complex competency questions identified.

3.5.2 Terminology Requirements

Creation of PTSDO specifications supports the refinement of user needs, requirements, and competency questions characterizing variations in symptomatology, and therapeutic intervention terms. **Table 3.6** provides a summary list of requirements gathered from recognized stakeholders. Each fall under one of three categories that includes requirements for design, modifications to the system, and for text mining support specification. For instance, REQ 1 is an important requirement outlining the design of PTSDO. It requires the reuse of identified existing concepts from vetted resources. In order to add specificity, completeness, and unambiguity, sub-requirements are attached to each higher-level prerequisite. A part of REQ 1, the system must be able to map to other terminologies (i.e. LOINC, SNOMED-CT, MFOEM). It is found to meet all quality criteria and is prioritized as essential. As a part of this requirements, each concept must include a textual or logical definition in order to identify a contextually comparable entity. At

minimum, an attached ‘example of usage’ will be acceptable inclusion until the definition is developed and incorporated.

ID	Requirement	Quality	Priority
REQ1	PTSDO shall reuse existing concepts and terms from available and reputable sources.	met	essential
REQ2	PTSDO shall be maintained to provide exports of all or part of the knowledge in standard XML formats.	met	essential
REQ3	PTSDO shall maintain traceability of the entities represented within the framework via Internationalized Resource Identifier (IRI).	met	essential
REQ4	Concept names must be unique, unambiguous, and provide context to its intended logical and definitional use.	unmet	essential
REQ5	Concepts shall be validated prior to input into PTSDO.	unmet	essential
REQ6	User shall not input experimental, trial, or off-label therapeutic intervention concepts.	met	conditional
REQ7	PTSDO shall retain semantic types of imported terms.	met	conditional
REQ8	PTSDO shall assist cTAKES in support of concept extraction.	met	conditional
REQ9	PTSDO shall be compliant with the W3C’s Web Ontology Language – OWL specifications.	met	essential
REQ10	PTSDO shall support linking and complementing existing resources that provide information about PTSD concepts.	unmet	conditional
REQ11	Each term within PTSDO shall be named using singular nouns.	unmet	conditional
REQ12	User shall not delete retired classes and instead relocate each to the deprecated class.	unmet	conditional

Table 3.6. Subset of terminology requirements

REQ 4 states that “concept names must be unique, unambiguous, and provide context to its intended logical and definitional use.” REQ 4 meets the qualitative criteria of completeness, consistency, traceability, relevancy, and unambiguity. With attached sub-requirements, consensus among stakeholders determined this requirement to be essential. Supportive and qualitative statements are attached to the requirement for specificity and consistency. For REQ 4, if the same term to be added is commonly used to mean different concepts in different contexts, then its name is explicitly qualified to resolve this ambiguity. If multiple terms are used to mean the same thing, one of the terms is identified as the preferred term and the other terms are listed as synonyms. REQ 8 defines the requirements of cTAKES in support of concept extraction which is conditional on the ability to maintain compatibility. PTSDO must be able to identify each PTSD symptom and treatment named entity. Many of the components are tuned for cTAKES compatibility, however the system is built to interoperate with several text mining

pipelines including MetaMap and NCBO Annotator. REQ 5 refers to a validation process to be maintained before entity inclusion into PTSDO. Entries are validated and vetted by management team or domain experts to determine domain necessity and to prevent duplicate entries. Necessary concepts for PTSD include those used in the diagnosis, screening, or prediction of PTSD. Examples of validated concept sources include resources such as DSM, APA, VA/DoD, or the National Institute of Mental Health.

3.5.3 PTSD Terminology System Development

The PTSD terminology system is being developed in order to improve the understanding of the disorder. A goal is to make assumptions explicit and reduce the ambiguity in concepts that describe symptoms and treatments within the domain. Its implementation will support the information extraction of narrative text located in the various sources within the biomedical domain. A goal of PTSDO is to enhance the knowledge of prediction, prognosis and recovery of patients diagnosed with this disorder by assisting clinicians and researchers with text mining initiatives.

3.5.3.1 Specification

PTSDO will be designed to support natural language processing (NLP) tasks, entity identification in narrative text, and the collection and sharing of PTSD information. The determined requirements exist to ensure these tasks adhere to standardization. The coverage is directed toward the categories: 1) variations in symptomatology; and 2) therapeutic interventions. The PTSDO concept coverage shall be accurately organized in order to capture terms the PTSD domain. Its scope is currently restricted to variations in symptomatology and therapeutic interventions. Symptom concepts related to PTSD include signs, symptoms, emotions, behaviors, mental processes, interpersonal processes, physiological responses, and mental dispositions. Treatment concepts related to the disorder include psychosocial training, psychotherapy, pharmacotherapy, case management, alternative medicine, and psycho educational training and materials. The arrangement will organize disparate data in order to enhance the clinical knowledge and understanding of the disorder. It must support overcoming the subjective nature of clinical PTSD diagnoses and symptom identification.

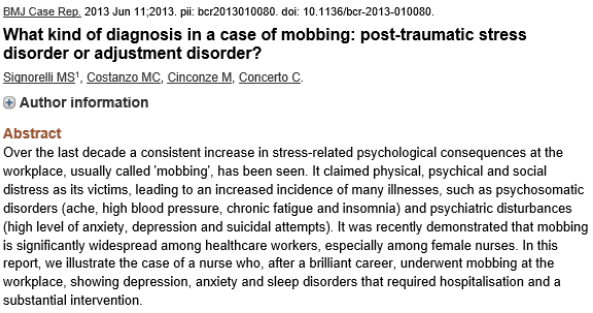
3.5.3.2 Manual Knowledge Acquisition

Manual term recognition is performed by relying on conceptual knowledge, i.e. human identification of terms and relating them to corresponding concepts. With consultation from domain experts, a base of concepts was gathered from SNOMED-CT which included terms, synonyms, and definitions applicable to the PTSD sub-domain as shown in **Table 3.8**. Clinical guidelines that were not machine-processable were also manually mined using linguistic analysis techniques to identify needed domain knowledge. This vocabulary must be focused in order to provide a framework for a foundation upon which to build NLP-based concept extraction. The concepts derived are added to the generic base of the knowledge representation to be more inclusive of available reusable resources without making assumptions that compromise its construction or usability. If the identified term is not in the dictionary, the previously discussed techniques of locating the term in an existing resource is conducted, if found it is incorporated into the terminology system, otherwise it is created de novo and incorporated into the terminology system. The text that identified the error is then re-processed through text mining pipeline. If the missing concept is recognized in the second iteration, the next corpus is analyzed, otherwise system and terminology error analysis is performed again until identification is achieved [198]. The terms gathered for the concepts of symptoms and treatments ranged from one word to a phrase including up to several words. Stakeholder interviews provided necessary but missing vocabulary from early terminology system prototypes. Coding applied to the transcripts of several cognitive interviews identified additional salient terms necessary to add for coverage. Next, annotations with symptom and treatment terms extracted from mental health notes of patients with PTSD obtained from the Veterans' Health Administration Corporate Data Warehouse (VHA CDW) was incorporated. Next, annotations with symptom and treatment terms extracted from mental health notes of patients with PTSD obtained from the Veterans' Health Administration Corporate Data Warehouse (VHA CDW) was incorporated. The CDW is the physical implementation of this logical data model at the enterprise level for VHA which consolidates data from disparate sources. CDW supports fully developed subject areas in its production environment as well as supporting rapid prototyping by extracting data directly from source systems with very minor data transformations [9, 34]. The contents of PTSDO after manual knowledge acquisition include concepts, synonyms, and their respective hierarchical organization which has been validated by clinicians and domain experts.

3.5.4 Automated Knowledge Acquisition

Because the manual term recognition approaches are time-consuming, labor-intensive and prone to error, the semi-automated methods described in **section 3.4.4** were implemented. The NLP text mining pipeline and semi-automated techniques were applied to electronically available reference documents, clinical guidelines. and downloaded literature related to PTSD. These resources are processed with the methods previously described to recognize candidate concepts to be added to the terminology. Processing continues until a satisfactory level saturation is achieved where the processing of documents is no longer acquiring new concepts not contained within the terminology system.

For example, a MEDLINE citation downloaded from PubMed is shown in **Figure 3.13**. These types of plain text files are processed via the cTAKES pipelines for concept mining. cTAKES produces annotations in XML output which includes the concept name, its part-of-speech, semantic type, and associated concept unique identifier (CUI). To aid the manual linguistic analysis, Python coding traversed the XML fields, extracted the annotated data, then outputted the findings within the plain text of the input. An example of this post-processing output is shown in **Figure 3.14**. Identified concepts are shown in red text and included meta-data annotation follows in parentheses.



BMJ Case Rep. 2013 Jun 11;2013. pii: bcr2013010080. doi: 10.1136/bcr-2013-010080.

What kind of diagnosis in a case of mobbing: post-traumatic stress disorder or adjustment disorder?

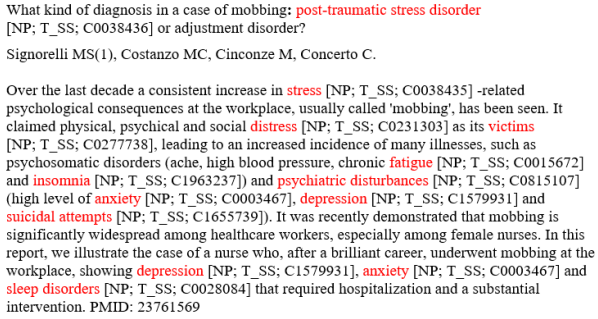
Signorelli MS[1], Costanzo MC, Cinconze M, Concerto C.

Author information

Abstract

Over the last decade a consistent increase in stress-related psychological consequences at the workplace, usually called 'mobbing', has been seen. It claimed physical, psychical and social distress as its victims, leading to an increased incidence of many illnesses, such as psychosomatic disorders (ache, high blood pressure, chronic fatigue and insomnia) and psychiatric disturbances (high level of anxiety, depression and suicidal attempts). It was recently demonstrated that mobbing is significantly widespread among healthcare workers, especially among female nurses. In this report, we illustrate the case of a nurse who, after a brilliant career, underwent mobbing at the workplace, showing depression, anxiety and sleep disorders that required hospitalisation and a substantial intervention.

Figure 3.11. Abstract retrieved for the query 'PTSD case report'



What kind of diagnosis in a case of mobbing: post-traumatic stress disorder [NP; T_SS; C0038436] or adjustment disorder?

Signorelli MS(1), Costanzo MC, Cinconze M, Concerto C.

Over the last decade a consistent increase in stress [NP; T_SS; C0038435] -related psychological consequences at the workplace, usually called 'mobbing', has been seen. It claimed physical, psychical and social distress [NP; T_SS; C0231303] as its victims [NP; T_SS; C0277738], leading to an increased incidence of many illnesses, such as psychosomatic disorders (ache, high blood pressure, chronic fatigue [NP; T_SS; C0015672] and insomnia [NP; T_SS; C1963237]) and psychiatric disturbances [NP; T_SS; C0815107] (high level of anxiety [NP; T_SS; C0003467], depression [NP; T_SS; C1579931] and suicidal attempts [NP; T_SS; C1655739]). It was recently demonstrated that mobbing is significantly widespread among healthcare workers, especially among female nurses. In this report, we illustrate the case of a nurse who, after a brilliant career, underwent mobbing at the workplace, showing depression [NP; T_SS; C1579931], anxiety [NP; T_SS; C0003467] and sleep disorders [NP; T_SS; C0028084] that required hospitalization and a substantial intervention. PMID: 23761569

Figure 3.12. Post-processing for streamlined concept/term identification

The ratio between knowledge being acquired and the number of abstracts being analyzed can be described via a saturation curve. The increase in concepts acquired and productivity of the analysis is visually displayed. PubMed clinical queries for PTSD (http://www.ncbi.nlm.nih.gov/pubmed/clinical?term=PTSD) were implemented in order to determine the corpus for text mining. A narrow scope query is implemented for the categories of

diagnosis, prognosis, and therapy. The query returned 2342 candidate articles for which a final analyses consisted of 219 full-text articles and 950 abstracts. Due to project time-frame, the mining of full-text articles is substituted with the mining of respective PTSD abstracts. In order to explore any concept-identification reduction, 5 percent of the 219 full-text articles were randomly selected for an ad-hoc analysis to detect if differences exist. There was no difference in those selected thus text mining continued using the PubMed abstracts.

The downloaded literature from PubMed is mined using the identified full-text articles and abstracts. The saturation curve for the concepts identified from semi-automatic linguistic analysis of PubMed citations is shown in **Figure 3.13**. Saturation was reached at around 700 documents. For the last 200 documents analyzed, zero new concepts or terms was identified. 93 reference documents were mined to return a total of 189 terms. Saturation was reached at around 70 documents. The saturation curve for clinical guidelines is shown in **Figure 3.14**. Clinical guidelines were analyzed on a per section basis for a total of 200 sections. Semi-automated mining reached saturation at around the 160th section. Utilizing methods from the Semantic Knowledge Representation group at NLM, a manual linguistic technique is employed in order to identify all concepts and terms that are missed by the system which can labeled as error [198]. The types and reasons for the errors vary greatly and consist of word sense disambiguation, missed negation, spelling, or missing from the PTSD dictionary to name some of the errors. The focus of this knowledge acquisition is the concepts and terms that do not exist in the PTSD terminology and must be added. Permitting for an inclusive interpretation, the analyses concludes that the curves describe the relationship between coverage accuracy of the terminology framework based upon the knowledge itself.

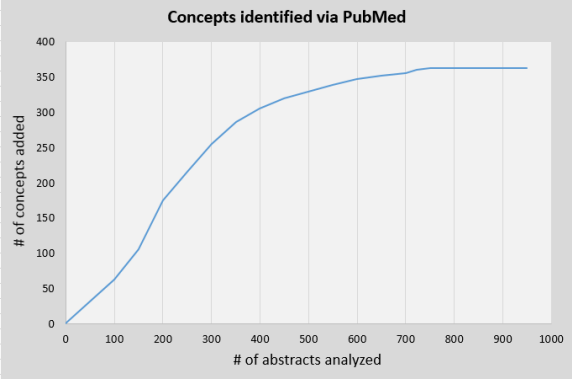


Figure 3.13. PubMed concept discovery saturation curve

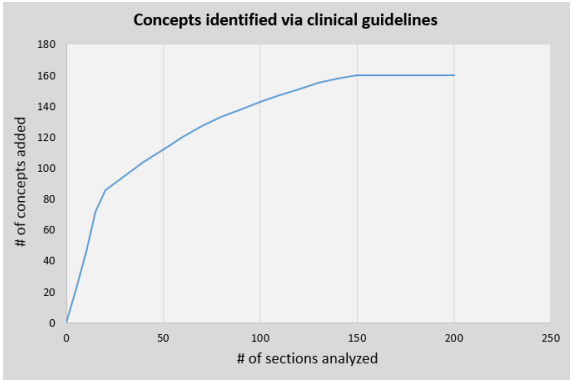


Figure 3.14. PTSD clinical guidelines concept discovery saturation curve

The data sources shown in in **Table 3.8** consist of the primary sources mined for vocabulary for input into the terminology system. Concepts acquired from clinical guidelines consisted of 62 symptom terms, and 98 treatment terms. For acquisition from SNOMED-CT, it was manually mined to obtain 172 symptom and 84 treatment terms. Reference document analysis obtained 102 symptom terms, and 87 treatment terms. 236 symptom and 127 treatment terms were acquired from the literature reviews. A total of six user interviews were conducted for knowledge acquisition. The number of terms gathered from these collections of experts were 37 symptom and 22 treatment terms. After mining the VA warehouse, a total of 2907 symptom candidate terms were collected and 1590 treatment candidates acquired.

Data Source	Symptom (n) terms	Treatment (n) terms
clinical guidelines	62	98
SNOMED-CT	172	84
reference documents	102	87
literature reviews	236	127
user interviews	37	22
VHA annotations	2907	1590
Total	**3516**	**2008**

Table 3.7. Collection of terms from various resources

The terminology system is modelled for five primary parent classes for symptom concepts within the structure. PTSD symptoms are arranged in clusters according to definitions in the Diagnostic and Statistical Manual of Mental Disorders [6]. Clusters, listed in **Table 3.9a**, include stressors, intrusion symptoms, avoidance and numbing, negative alterations in cognitions and mood, alterations in arousal and reactivity, functional significance and dissociative symptoms. It is important to semantically distinguish these variations in symptoms as they translate directly into the diagnosis of disease and type and breadth of clinical care. While the variations in symptoms are applicable to multiple cohorts, the context of this framework is derived from adult patients with traumatic stress reaction treated in a Veterans Healthcare Administration (VHA) clinical setting. This symptom grouping establishes parameters necessary for the semantic understanding of assessment, diagnosis, and management of symptoms. Similarly, concepts describing treatment interventions are arranged in categories designated in the Veteran

Affairs/Department of Defense (VA/DoD) PTSD evidence-based practice management guidelines [34].

PTSDO is modelled for five parent classes for treatment concepts. Five primary therapeutic categories of pharmacotherapy, psychological, psycho educational, psychosocial, and case management are shown in **Table 3.9b**. Knowledge about the variations in available prescribed treatments for PTSD can further enhance the ability to comparatively evaluate their relationships and effectiveness on treating the symptoms of this disorder. This organization provides structure for symptom-specific management supporting precision of classification. The hierarchical arrangement of symptoms and treatments allows representation of data using parent/child relationships and fosters organization of information. As the gaps in current understanding of the disorder are addressed, it is important for the structure to set parameters that foster contextual collaboration.

Parent Class	Subclass (N =)	Synonyms (N =)
Trauma Exposure	118	92
Re-experiencing	82	81
Avoidance	352	552
Elevated arousal	242	377
Functional impairment	99	87
Total	**893**	**1189**

Table 3.8a. Symptom classes and synonyms

Parent Class	Subclass (N =)	Synonyms (N =)
Psychological	191	52
Psychosocial	11	4
Psycho-education	36	5
Case management	54	14
Pharmacologic	59	74
Total	**351**	**149**

Table 3.8b. Treatment classes and synonyms

PTSDO consists of taxonomic relationships which organizes its concepts into a sub- super-concept tree structure. This framework utilizes the "is-a" relationship to form a subsumption taxonomy between each subclass with its supper class. Selected symptom taxonomic relationships for the framework are shown in **Figure 3.15** with several subclass concepts listed. For example, the concept C1821940: flashback is a subclass of the cluster re-experiencing. Its required meta-data within the framework includes the following:

label: flashback

preferredLabel: flashback episode

IRI: http://purl.bioontology.org/ontology/SNOMEDCT/30871003

Semantic Type: Finding

definition: a strong memory of a past event that comes suddenly into one's mind

definition source: http://www.meriam-webster.com/dictionary/flashback

example of usage: flashback of traumatic experience

example of usage: feeling as if the traumatic event were recurring

hasExactSynonym: recurring thought of traumatic experience

hasExactSynonym: reliving experience

The PTSDO terminology framework is task-oriented toward the support of text mining annotation analysis. It is focused on terms and concepts specifically describing variations in symptomatology and the diverse therapeutic interventions that persist. New concepts, nomenclature, and changes to information need to be incorporated into the terminology system as research advances and findings evolve. For de novo terms input, complete definitions and examples of usage should be documented. For example, ‘act out’ is a de novo term with no available terminology resource that presented a contextually accurate definition. This textual string is a salient arousal concept commonly used to describe patient actions and must be captured. The stakeholders via committee developed the following textual definition from various online dictionary resources: “to behave badly or in a socially unacceptable often self-defeating manner especially as a means of venting painful emotions (as fear or frustration).”

Figure 3.15. Selected symptom taxonomic relationships for PTSDO

Figure 3.16 displays selected treatment classes and specific subclasses of the psychological treatment concept. For example, the concept C0015618: family therapy is a subclass of the cluster psychological (therapeutic intervention). Its required meta-data includes the following:

label: family therapy

preferredLabel: Family psychotherapy

IRI: http://ncicb.nci.nih.gov/xml/ow/EVS/Thesaurus.owl#C93347

Semantic Type: Therapeutic or Preventive Procedure

definition: type of therapy in which the whole family talks with a professional counselor to solve family problems

definition source: NCI/NCI-GLOSSPT

example of usage: engaged in therapy with family

hasExactSynonym: family psychotherapy

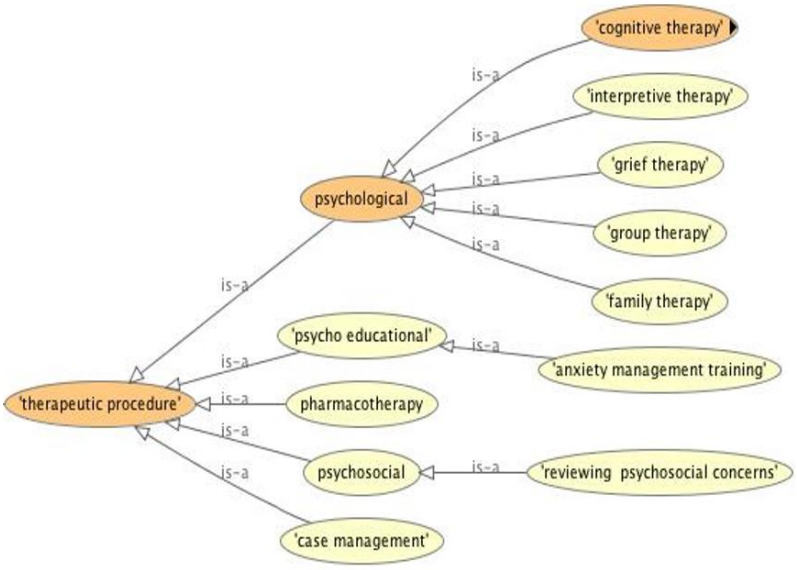


Figure 3.16. Selected psychological treatment classes

3.5.5 Concept and Term Overlap

PTSDO is integrated using and re-using several accessible general and domain specific terminologies. This reuse of well-defined concepts supports effective sharing of the concepts and terms within the terminology framework. Each entity identified in an outside source relevant to

be added to PTSDO is linked to the original terminology resource with its Internationalized Resource Identifier (IRI).

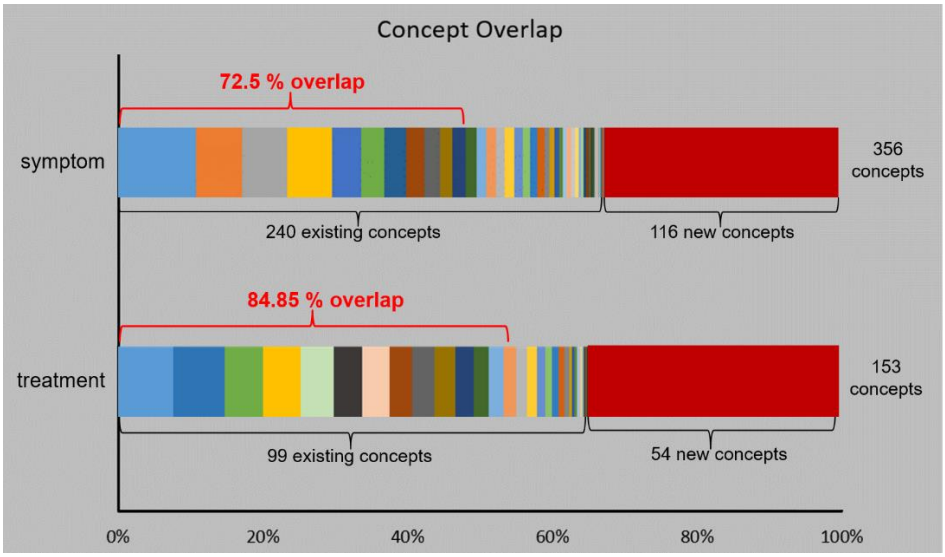


Figure 3.17. Symptom and treatment concept overlap of PTSDO

The quantity of symptom and treatment concepts within PTSDO is displayed in **Figure 3.17**. 240 (67.42%) of symptom concepts exists in one or more available terminology resources out of the 356 total symptom concepts in PTSDO. These concepts were imported and externally referenced to its respective source terminology via IRI. For example, the concept of 'rage' is imported from the Emotion Ontology (MFOEM) and linked to http://purl.obolibrary.org/obo/MFOEM_000014, the source concept identification. Each update to rage upon import is updated into PTSDO to include its metadata. 116 symptom concepts were not available in an external resource thus having to be created de novo. While the concepts of C00036558: self and C0870209: blame are available for post coordination in existing resources, the singular entity self-blame is not. This concept is created de novo and by stakeholder committee given the definition "to have a perception, emotions, or thoughts that oneself is responsible for something bad that has happened." 99 (64.7%) of treatment concepts are found in existing resources which necessitated 54 de novo created. An example of a de novo treatment concept is stress inoculation training. Although types of stress inoculation are available, the concept itself is not. An imported treatment concept is http://ncicb.nci.nih.gov/EVS/Thesaurus.owl#C15986, C0013216: pharmacotherapy.

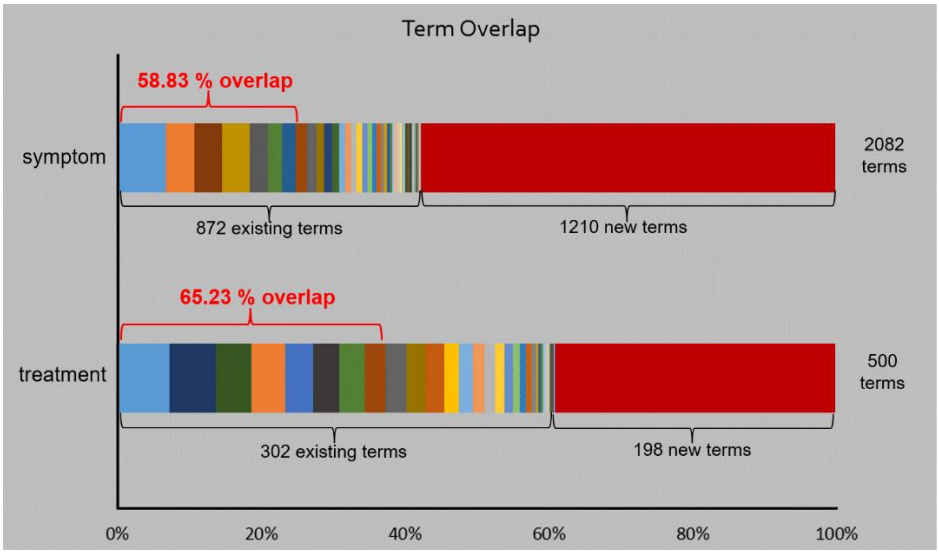


Figure 3.18. Term overlap

Term quantity is shown in **Figure 3.18** displaying the collection of symptom and treatment terms acquired and inputted into PTSDO. A count of terminology system terms is separated by subclasses and synonyms. PTSDO contains a total of 2582 terms, of which 1174 (45.47%) terms were identified in existing resources. A total of 1408 (54.53%) terms were created de novo. There are 1244 subclasses identified for PTSDO, of which 735 (59.08%) terms were identified in existing resources and 509 (40.92%) new subclasses created.

Of the 2082 symptom terms in PTSDO, 872 (41.88%) terms were identified in existing resources and 1210 (58.12%) terms were created de novo. There are 893 symptomatology subclasses in PTSDO. Subclass terms includes inherited classes from the parent DSM categories such as suicide plan, closed off, and paranoid thought. 477 (53.42%) of the subclasses was identified in existing resources and 416 (46.58%) were created de novo. These subclasses include various synonyms and acronyms that have an exact synonymy. For example, C0424366: self-harm contains exact synonyms of self-injury, self-damage, self-poisoning and acronyms of SI, and DSH. While self-harm does have several synonyms, it is defined within the meta-data of the PTSDO in order to establish an accepted meaning with this medical sub-domain. Examples of symptom de novo terms created are torn, on edge, and difficulty with authority. 395 symptom synonyms (33.22%) were found in existing resources and 794 (66.78) de novo synonyms were created.

Treatment terms in PTSDO totaled 500. Existing resources identified 302 (60.4%) terms and 198 (39.6%) de novo terms. Of the 351 subclasses, 258 (73.5%) were found in existing resources and 93 (26.5%) were created. Parent classes for treatment subclasses are developed from the VA/DoD clinical practice guideline for post-traumatic stress management [34]. Examples of a class C0009244: cognitive behavior therapy contains relevant subclasses that include C0966644: cognitive processing therapy, C0870527: exposure therapy, and C0582604: thought-stopping training. Existing treatment synonym terms totaled 44 (29.53%) and 105 (70.47%) terms created de novo of the total 149 synonyms. Treatment synonyms included abdominal breathing, deep breathing, and belly breathing for subclass C0231898: diaphragmatic breathing. Acronyms for C0009244: cognitive behavior therapy includes CT, and CBT.

The reuse of terminology either by Internationalized Resource Identifier (IRI), explicit references (*xref*), or a Concept Unique Identifier (CUI) mapping within existing resources averages to just slightly more than 8%. 92% of exact terminology exists in multiple terminologies but do not reference another source. The only resources to reuse concepts are from ontologies with membership to the OBO Foundry. Out of a total of 94 sources that contained PTSDO concepts and terms, only 8 resources mapped each to other terminologies via either an IRI construct, explicit reference, or CUI. 86 did not map any of its classes at all. The largest quantity of symptom PTSDO classes are from SNOMED-CT, MEDDRA, NCIT, APAONTO, and RCD-CT that each have more than 40 PTSD entities available for reuse. The highest quantity of therapeutic intervention PTSDO classes are from SNOMED-CT, MESH, RCD-CT, APAONTO, RXNORM, and LOINC each have more than 30 PTSD entities available for reuse.

Average inclusive concept overlap for PTSDO is 78.58, while term overlap is 62.03% averaged over the other resources. 174 of the 240 (72.5%) existing symptom concepts are found in two or more terminologies with no external reference. Existence in one more sources with no external reference for symptom terms is 513 out of the 872 (58.83%). C0001818: agoraphobia is represented in 20 available terminology resources with no external reference to other resources. C0015672: fatigue is not externally linked and found in 17 sources. Of the existing treatment concepts, 84 of the 99 (84.85%) is found in more than one terminology with no external reference. 197 of the 302 (65.23%) existing treatment terms is found in more than one terminology with no external reference. A large percentage of the treatment overlap is from medicines of which over 34 occurs in 10 or more sources. For example, C0002600: amitriptyline

a subclass of C0003290: tricyclic antidepressant is found in 19 of the terminology resources. Aside from pharmacotherapy, C0085971: case management is represented in 7 of the terminology resources.

3.6 Discussion

Manual mining of existing terminology resources performed during needs assessment produced a generic PTSD terminology system base aiding in the maximization of stakeholder input. Manual knowledge acquisition of existing terminology resources did not produce the comprehensive needed concept or term coverage. The iterative semi-automated text mining approach to knowledge acquisition successfully determined sufficient coverage to satisfy identified stakeholder needs. Concept and term overlap is found to exist by more than 50% in identified existing available resources. Reuse of entities is very low and only adhered to by select ontologies.

Developing requirements are commonly one of the greatest challenges in terminology development regardless of the methodology utilized. There are certain issues that come about when discussing the development and usage of terminology structures. Across all stakeholders, there is a high prevalence of lack of understanding regarding the vocabulary structures definitions and what is offered by domain terminology development. Although stakeholder interaction was at times extremely limited in this research, allocating time for presentation of structure and framework capabilities was valuable use of meetings. This aided in reduction of misunderstandings between stakeholders. Focus group sessions are difficult to coordinate with the busy and conflicting schedules of stakeholders. However, the sessions did provide the most useful feedback when organized appropriately. The interaction of the stakeholders fosters mutual understanding for the value of developing a PTSD terminology. Synergistic ideas and input is obtained from their respective collaboration. This alliance of users is paramount to impacting the evolution of the PTSD conceptual framework. Group sessions foster interactive feedback, promote involvement, and provide education to all involved. It cannot be stressed enough how important it was for the project to locate a clinician champion, as substantial progress was not made until their support was given. This fostered stakeholder understanding of the project that allowed engagement of sympathetic participants.

The importance of the requirements gathering process was not fully realized until mid-project when early errors in the specification process trickled down and impacted PTSDO design and its implementation into text mining pipelines. A benefit of implementing agile methodology techniques allowed for these type of errors to be identified and corrected quickly while not impacting the momentum of the entire project. The iterative development provided with agile methods was critical for requirements gathering considering the lack of clear ideas of system capabilities. There was a definitive understanding gap of what terminology and text mining systems could provide to users. There were several iterations of revising needs and requirements that slowed terminology development from the traditional gathering methods. Maintaining a working prototype allowed knowledge acquisition to continue while stakeholders debated requirement prioritization. Many of the user needs conflicted with other stakeholders, and development would have experienced many stalls without use of the agile method. The majority of the terminology system requirements identified addressed terminology system design, concept inclusion, and text-mining interoperability support. The terminology is designed with a documentation process that promotes agile maintenance and change processes. These requirements maximize the terminology's overall usability and encourages its support of unpredictable future implementations of the system. Imprecise needs were the biggest challenge as development issues arose when requirements were not precisely stated and ambiguous descriptions varied interpretation. The documentation of changes is imperative to prevent the repetition of mistakes which had the likelihood of occurrence due to the many stakeholders.

There were some organizational and political factors that influenced the final modeling of PTSD domain knowledge. Modelling symptoms according to DSM arrangement and treatments corresponding to VA/DoD guidelines was mandated despite stakeholder acknowledgement that this would inhibit community acceptance outside of the Veterans Healthcare system. As new stakeholders were brought on throughout development, there were many changes that had to be implemented which were difficult even with agile method applications. The process of requirements elicitation itself generates more detailed and creative thinking about the problem of domain coverage that in turn affected the scope. As the possibilities for a solution emerge, there are numerous decision points concerning what should and should not be included within the scope of PTSDO. As knowledge acquisition began, the collection of terms overwhelmed stakeholders leading to the limiting of coverage to only symptoms and treatments for the initial

design. Ideas were lofty at the outset of the project and attempts to model too much of the PTSD domain was quickly identified as not feasible. The coverage of the framework currently includes only the most salient vocabulary for text mining tasks.

The use of techniques acquired from the agile methodology is found to be valuable in the development of the framework. Since the agile methodology is not specifically designed for terminology development, its implementation is not without difficulties. However, its usefulness can be summarized by its style of recognizing the problem, fixing it, and continuing on to the next iteration with its set of issues. The identification of elements that are useful for specific tasks is particularly challenging. Initial design steps were a trial by error to discover those agile techniques that were not practical. As beneficial as the agile approach is to expedite and development adaptability, it is not an approach that maximized reusability. Domain modelling complications arose due to the many comorbid mental health symptoms that were collected during knowledge acquisition. Many psychiatric disorders have very similar symptoms and treatments with conditions such as substance abuse, and other anxiety disorders. It is difficult to separate PTSD symptoms and treatments from several other mental health disorders. The pace of development often alternated between rapid achievements and slow incremental task completions. Due iterative development, it is necessary to continuously switch from one stage to another within the design methodology framework implemented. Stakeholders change their minds for many reasons, and do so on a regular basis. A formidable challenge that occurred while developing the terminology framework was the concurrent documentation management and knowledge acquisition. Utilizing the developed framework to mine concepts to subsequently add to the terminology is a time consuming process inhibited by project management responsibilities and documentation requirements. There are also challenges in concept and term selection from existing resources. Many of the automated tools designed to aid in this process are not user friendly, provide incomplete information, or not interoperable with all relevant terminology resources. Even in accessible sources, many of the terms are not well documented thus the context is difficult to understand. Also, many definitions of existing terms are missing while others are ambiguous in their intended meaning. Attaching meta-data is time-consuming and challenging due to difficulty in acquiring the necessary information. PTSDO contains no axioms to limit what knowledge is capable of being captured with text-mining support nor specific individual instances. This design is purposeful as text-mining support minimally requires

dictionary-lookup support. Future implementations of PTSDO will include the axioms, instances, and formal logic within the terminology.

For missing definitions, development of an unambiguous textual description requires much deliberation, rigorous examination, and stress-testing. Each vocabulary term within PTSDO is associated with a rigorous definition. Several stakeholder meeting sessions were encumbered by this unforeseen challenge of definition development. Those created for de novo terms in the terminology strive for precision with as few elements as possible. Many of the first iterations for terminology descriptions were overly vague leaving the possibilities of misinterpretation. The included vocabulary terms require granularity in order to maximize recall of PTSD terminology text mining initiatives.

There is a prevalent issue with concept and term overlap among knowledge acquired from existing resources. The reuse of terminology despite exact conceptual usage similarity is extremely inadequate among existing resources. Term reuse with appropriate external reference is primarily limited to entities belonging to ontologies with OBO Foundry [128] membership. For an ontology to vetted by the OBO Foundry, it must adhere to a set of principles for orthogonal interoperable reference ontologies [128]. Reuse such as in the mental functioning ontology (MF) [14] achieves orthogonality to be exemplified. It reuses 91.33% of its terms from 6 different ontologies providing an excellent template in the development towards a PTSD ontology. The UMLS achieves excellent external references via concept unique identifiers (CUIs) but importing and linking the data it contains is extremely difficult. There is a rigorous process for terminology to be added to be assigned a CUI perpetuating independently developed vocabularies to not add respective identifiers for linking synonymous terms. Although the UMLS is comprehensive, it does not fit all use cases of applications requiring terminology support. Many researchers, ontologists, or terminologists do not reference the system in order to locate the appropriate concept.

The majority of medications inputted into PTSDO are available in 10 or more sources, however none are linked to one another or reused. SNOMED-CT contains more concepts and terms deemed necessary for PTSD domain modelling than any other available resource. Below SNOMED-CT, the majority of symptom coverage is achieved with MEDDRA and the NCI Thesaurus. Additionally, MESH terms and the APA psychology ontology contain a substantial amount of the treatments included. The value of terminology reuse needs more attention in the

development community. Term reuse within PTSDO can reduce engineering costs and support semantic interoperability among different datasets and applications. The great amount of overlapping terms with minimal reuse in comparison goes against the purpose of terminology standardization, interoperability, and sharing that needs to be pursued much more aggressively in development initiatives.

3.7 Conclusion

The requirements analysis enabled users and designers to compile agreed upon functional needs and desires that are to be met by the framework. The development for a PTSD terminology framework has set the groundwork for future enhancements of the vocabulary to answer new and expanded competency questions. PTSDO provides structure to information, enables reuse of the knowledge base, and removes concept ambiguity. The subjective nature of clinical PTSD diagnoses and symptom identification in the narrative nature of biomedical text is a constraint the terminology system strives to address. The collection of the domain knowledge contains traceability of resources which promotes sharing of information in various healthcare settings. The expansion of PTSDO towards an ontology will enhance interoperability and allow much more sophisticated use case support. Future development of the terminology system with linkage of data will enable system designers to identify data sources used to derive the ontology concepts. Changes can then easily be made to fit local variations or to update ongoing evaluations.

Natural terms that are commonly used within the PTSD research community are modeled within the framework. When term additions introduced ambiguity, alternative terms are explored. Consensus of meaning and context is acquired from stakeholders to support term interpretation and reduce errors. The developed framework in this research organizes domain information with formalized structure that facilitates improved text mining. This is portrayed in the dictionary support of concept discovery used in the semi-automated knowledge acquisition. The scalable design of PTSDO allows for ease of additions for increased future coverage, refinements to scope, and modelling revisions.

This research project has provided a common vocabulary for clinicians, researchers, and developers who need to share information about PTSD. The PTSDO and terminological systems of the like provide the backbone to create powerful biomedical applications. They are crucial for

providing context and structure to the discovery of information. Long-term healthcare and life science research will benefit from quality of data and value to all stakeholders. The synthesized collection of domain terminology presented contributes to the expansion of the understanding of PTSD. Its continued development (https://github.com/bt29gamble/PTSD-terminology) towards an ontology will support overcoming the difficulty in describing the range of symptoms along with the wide array of possible treatments for the disorder.

Chapter 4

Implementation and validation of a PTSD dictionary for supporting text mining pipeline entity identification tasks

4.1 Introduction

The medical domain consists of a vast amount of narrative text ranging from published literature to clinical progress notes to online health forums. Biomedical text can consist of natural language text from journal articles, books, pamphlets or posters. In clinical progress reports, narrative free text is written by clinicians discussing patient encounters. This free text is convenient for expressing terms and events but it is difficult for secondary use such as searching or performing analysis required by many biomedical researchers. With the growing information overload in electronic medical records and biomedical literature, there is an increasing need for annotation in order to support natural language processing (NLP) [18] systems and information extraction (IE) [304]. Narrative text is continuing to accumulate and new ways must be found to access the information that it contains more efficiently [305-306].

Information overload is a well-researched problem of the last several decades but its pace of expansion has increased exponentially [307]. This is due to many factors including the expansion of research and testing, the decreasing costs of data storage, and the increasing sophistication of information delivery methods [308-309]. The development of applications to provide assistance has increased as well, however their respective success is variable. Key to managing information overload in the biomedical domain is the focus on specific knowledge needs. Developing and maintaining terminology sources with fully inclusive coverage of a sub-domain provides great potential of supporting these emerging applications to impact the problems of information overload by improving accuracy.

NLP provides an interesting perspective on text mining application as it focuses on computational linguistics and the interactions between computers and human language. Tan [310] refers to text mining as a discovery process for “extracting interesting and non-trivial patterns or knowledge from unstructured text,” textual databases [311], and relevant documents

[310]. Text mining has been a vetted research application to apply beneficial usability to this compounding problem of information overload. Cohen and Hersh differentiate between text mining and NLP stating that NLP "attempts to understand the meaning of text as a whole, while text mining and knowledge extraction concentrate on solving a specific problem" [16]. NLP began in the 1950s as the intersection of artificial intelligence and linguistics [312]. The importance of well-developed NLP arose out of the limitations of relying purely on rule-based systems. The needs of biomedical applications have exponentially advanced making necessary rules unmanageable. Natural language is vast in size, complexity and ambiguity which led to emphasis on using statistical-based approaches in order to extract meaning from text [25]. By making use of large amounts of data, statistical approaches achieve meaningful results and great progress through active learning [313]. However, with unconventional data that all researchers, scientists, and clinicians must anticipate, these highly accurate statistical-based approaches to NLP will degrade [25].

Information extraction (IE), sometimes described an application of NLP, supports domain extraction by structuring desired information from narrative text while ignoring irrelevant information [314]. IE systems foster knowledge engineering for processing corpora and machine learning using patterns to extract data [163, 315]. Named entity recognition (NER) is one of the most common uses of information extraction technology [316]. NER identifies terms or phrases and categorizes them according to specified entities. A common NER task is mapping named entities to concepts in a vocabulary [25]. The ability to recognize previously unknown entities centers upon recognition and classification rules triggered by distinctive features associated with training examples. The supervised machine learning approach typically consists of a system that reads a large annotated corpus, memorizes lists of entities, and creates disambiguation rules based on discriminative features. A semi-supervised learning approach uses bootstrapping by starting with a set of references for the learning process of the system [317]. Even greater results can be achieved via NER supported with fully developed terminology resources. These resources can take many forms including a list, dictionary, gazetteer, lexicon, ontology, taxonomy, thesaurus, or controlled vocabulary. Some of these resources provide means to express the relation "is a" (e.g., *fear is a symptom*). If a word (*fear*) is an element of a list of symptoms, then the probability of this word to be symptom in the given text is high. However, because of word polysemy, the probability is almost never absolute [317].

Some NLP systems chain the analytical tasks as shown in **Figure 4.1** into a data flow pipeline. With this text mining pipeline, different techniques can be used for each task allowing a modular approach to the system design. The output of an analytical module becomes the input to its preceding module task. This design allows focus on improvements to be made at the task-level which increases the robustness of the NLP system as a whole.

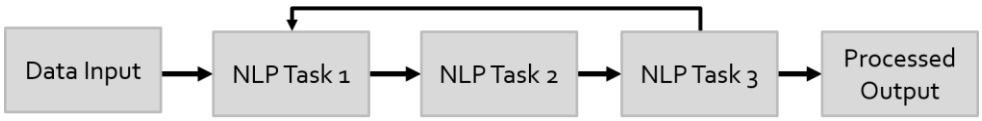


Figure 4.1. Modular text mining pipeline for NLP tasks.

NLP systems require a dictionary reference for mapping to concepts, interpreting meanings, and understanding interactions between words. This reference also provides information about part of speech, lexical functions, idioms, subcategorization and semantics [18, 25]. These NLP systems, whether using a lookup, rule-based, or statistical approach, require some degree of support with a dictionary or terminology resource [66, 147, 170]. A dictionary-lookup approach mimics a matching pattern technique and heavily relies upon the quality of the reference. Even the rule-based and statistical-based methods that reference an incomplete terminology, in essence, build a form of the dictionary within the system [18, 166]. Regardless of the selected approach, the chosen system has to know the lexicon being searched or lexical analysis needed whether built into an outside dictionary or embedded into the coding or rules. However, it can be much more difficult to maintain updates and changes within the NLP system or programming code when compared to a free-standing terminology resource [166, 182].

System design dictates the level of terminology coverage required so it is important to identify the appropriate text mining system for an intended use case. PTSDO is designed to support natural language processing (NLP) tasks, entity identification in narrative text, and the collection and sharing of PTSD information. The goal is to support tasks related to annotation and concept identification that adhere to standardization. Dictionary resources at a minimum must provide entity type name lists and classification schemas for NER. The focus on NER is a first crucial step in extracting more complex types of information [58]. Determining the level of entities or concepts within a terminology provides a means to evaluate its coverage. Building dictionary-based tools provide focused terminology support for clinical information systems to

improve accessible knowledge to researchers and clinicians [16]. Such dictionaries contain lexical vocabulary that enable the recognition of desired domain-specific entities. The design of a vocabulary computational model must support a machine-readable, syntactic and semantic focused lexicon made explicit and available for reuse to support information extraction. Without focus knowledge bases with sufficient domain coverage, it is difficult for biomedical applications and NLP systems to retrieve information from heterogeneous and autonomous electronic resources. It is important for information systems to use the terminology system in order to provide meaning and context into the NLP [196]. This extends the impact concept classification has on data quality of text mining and NLP tasks. Without an explicit common language, vocabulary accuracy and sufficient coverage will suffer from limited domain compliance [318]. There is a documented need for terms to be gathered for storage and retrieval which will ultimately foster sharing and interoperability of information [255, 280].

The range of textual understanding achievable by medical NLP and NER systems is often bounded by rather limited domain knowledge available from available terminology sources. Well-designed systems can exploit this informational knowledge to support a range of healthcare decisions. The accuracy of these decisions is proportional to the accuracy of the concept recognition systems [317, 331]. Determining this accuracy requires evaluation with high quality gold standards for a specific domain or use case. In many medical sub-domains, gold standards for text processing is inaccessible or non-existent. Creating gold standards can be difficult and costly to create because they require manual annotations of instances to support specific textual identification tasks. However, they are necessary in order to evaluate and understand the usefulness of an NLP or text mining system [326, 328]. Developing annotated corpora formatted with standards also can assist biomedical researchers and lead to automated support of acquiring knowledge from this narrative text [4, 305-306].

4.2 Background

4.2.1 Text Mining and Natural language processing (NLP)

Text mining is the discovery of previously unknown information from natural language, typically acquired from extracting patterns from natural language of unstructured textual resources [16, 151]. Natural language processing (NLP) can be defined as the set of approaches, techniques, and mechanisms used to analyze text corpora written in natural language [163, 319].

Approaches to NLP include lexicon-based, rules-based, and statistical though many systems permit a combination of approaches [16]. As shown in **Table 4.1**, Nadkarni et al. has stated that NLP tasks can be sub-divided into low-level and high-level tasks. Low-level NLP tasks include sentence boundary detection, tokenization, part-of-speech assignment (POS tagging), morphological decomposition of compound words, shallow parsing (chunking), and problem-specific segmentation. Most NLP systems perform all of the low-level tasks identified in **Table 4.1**. Typically, each of the low-level tasks must be completed before beginning any of the high-level tasks [25].

Nadkarni et al. identifies high-level NLP tasks that are usually problem-specific. The first is spelling/grammatical error identification and recovery. The second is named entity recognition (NER) which is the identification and categorization of specific words or phrases. Next, word sense disambiguation (WSD) determines a homograph's correct meaning. Then, negation with uncertainty identification infers whether a named entity is present or absent. Another high-level subtask is relationship extraction where relationships between entities or events is determined. Next, temporality makes inferences from temporal expressions and temporal relations. Lastly, information extraction (IE) identifies problem-specific information and its transforms it into a structured form [25].

Low-level tasks	**Description**
Sentence boundary detection	identifies complications from tokens such as periods in abbreviations and titles
Tokenization	identifies tokens within a sentence
Part-of-speech tagging	marking word part-of-speech assignments
Morphological decomposition	lemmatization; decomposition of compound words
Shallow parsing	chunking; identifies sections of text
Problem-specific segmentation	segments text into meaningful groups
Higher-level tasks	**Description**
Grammatical error identification	identifies spelling errors
Named entity recognition (NER)	identifies and categorizes specific words/phrases
Word sense disambiguation (WSD)	determines a homograph's correct meaning
Negation	determines presence or absence of an entity
Relationship extraction	determines relationships between entities or events
Temporality	makes inferences from temporal expressions and relations
Information extraction (IE)	transforms problem-specific information into structured form

Table 4.1. Low-level and high-level NLP tasks [25]

IE can be defined as an application of NLP with the goal of extracting information from a corpus [163] typically occurring in two steps. The first step uses low-level NLP tasks and, in the second step, applies a series of techniques and approaches to extract information. Although these steps both have their peculiarities, in many implementations they may be joined into one single step [163, 319]. The main strategies for IE on the biomedical domain, according to Ananiadou and McNaught, are dictionary approaches, rule-based, and statistical-based methods. A dictionary-based approach uses resources from terminologies to extract concepts. Rule-based approaches can use patterns to find and extract relevant concepts. Machine learning techniques can identify useful concepts in the corpus by previously annotated data. Lastly, the authors recognize that statistics-based approaches can calculate distributions of text inside a corpus in order to identify and extract information. The combination of dictionary, rule-based, and machine learning approaches can create an ensemble for achieving the best accuracy [163, 319].

NER involves identifying the boundaries of text, then mapping the entity to a unique concept identifier in a terminology resource in order to disambiguate and apply meaning [320]. The process labels sequences of words in a text which are names of things, such as a person, gene names, signs and symptoms, and most entities that can be categorized. It is also known as entity identification, entity chunking and entity extraction. It is a subtask of IE that seeks to locate and classify elements in text into pre-defined categories. NER systems have been created using linguistic grammar-based techniques as well as statistical models and machine learning approaches [317, 320].

4.2.2 Text mining pipelines

Text mining explores automatic or semi-automatic discovery of information from vast unstructured text [145]. As was shown in **Figure 4.1**, pipelines allow for the various sub-tasks to be broken down into modules where focus can be directed toward improving on of the specific tasks involved within the complete text mining system. This is the intention behind pipelined NLP frameworks, such as GATE [321] and Unstructured Information Management Architecture (UIMA) [322]. GATE provides a common infrastructure for performing NLP and IE tasks and is very popular in the text mining community because of its prevalent machine learning plugins [171]. UIMA is comparably as popular due to its scope that goes beyond NLP. The framework has the ability to integrate structured-format databases, images, multi-media, and any arbitrary

technology into the pipeline for analysis. In UIMA, each analytical task transforms a copy of its input by adding XML-based markup and/or reading/writing external data. A task operates on Common Analysis Structure (CAS), which contains the data, a schema describing the details of the markup formats, the analysis results, and links indexes to the portions of the source data to which they refer. UIMA utilizes the CAS to interact with the UIMA pipeline to oversee analytical tasks. This feature allows, within limits, each module or task to be written in varying programming languages [169, 180, 322].

Achieving perfect accuracy in a given NLP task is typically impossible as errors that manifest in one module will propagate to the next degrading accuracy at each task. Manifestations of errors is a problem for any NLP system regardless of the framework implemented. Developing NLP tasks through multiple pipelines has allowed developers to focus on each component individually with only minimal effect on subsequent modules. This method also fosters the tuning of recall and precision based upon user needs [25].

While a detailed discussion of the model details is beyond the scope of this dissertation, there are several machine-learning approaches to NLP problem solving. Statistical and machine learning involve development and implementations of algorithms that allow a program to infer patterns from training data, that in turn allows generalized predictions about new data [25]. These approaches can be applied to broad NLP problems or specific tasks of NER. Each approach is enhanced by the dictionary or list lookup features the system accesses for reference. List lookup features can consist of a general list of dictionary words, stop words, capitalized nouns, and common abbreviations. A list could also consist of a collection of entities such as findings, organizations, healthcare activities, or organism types similar to the list of semantic types utilized in the UMLS. Lastly, a list could be made up of entity cue such as typical words expected in the data, titles, or locations of typical words. List lookup features can permit fuzzy-matching that allows capturing of lexical variations in words [317].

The National Library of Medicine (NLM) provides several well-known ‘knowledge infrastructure’ resources that apply to multiple NLP-based tasks. The UMLS Metathesaurus, [197] which records synonyms and categories of biomedical concepts from numerous biomedical terminologies, is useful in clinical NER. The NLM’s Specialist Lexicon [323], accompanied by a set of NLP tools, is a database of common English and medical terms that includes part-of-speech and inflection data [324]. The NLM also provides a test collection for word

disambiguation [25, 196]. Often, the NER systems are only as good as the annotation that supports its intended tasks [146, 325]. NER systems are typically evaluated based on how their output compares with the output of human linguists known as manual annotation [317].

4.2.3 Text Annotation

Annotation involves identifying a target of a word or span of words and attaching attributes and descriptions to this intended target. As a methodology, annotation provides meta-data, which is data about a word, sentence, section or entire document. It is often done to assist in the acquisition of broader knowledge from the text as well as speed up information extraction. These annotations can be stored in a knowledge base, ontology or a controlled vocabulary. Manual annotation requires humans to identify specific terms and arrange them in a structure to be accessed. Many steps of the annotation process can be automated. Automation reduces a human annotator's responsibility of term identification as well as reduces the number of errors. However, manual annotation tasks are still required to build and train the automation. Automation can sometimes miss valuable information which is why many researchers use semi-automated methods supported by some form of dictionary. The dictionary or terminology resource can guide annotators that are not domain experts to better understand the intended hierarchy of arrangement of concepts. Use of an explicit terminology resource can increase comprehensiveness of each concept identification within the text [326-327].

Supporting NLP or machine learning tasks with annotated data fosters extraction specialized for the intended domain and the corpus aids in automation of discovery [328]. Semantic annotation applies meaning to annotation by reducing ambiguity associated with the text. It provides enrichment and context to the unstructured data by linking it to a structured knowledge base. The annotation connects a word or span of text to a terminology which transforms the terms into an entity [326]. Annotation of text can sometimes be mapped to terms from a controlled vocabulary [329], such as those contained in the UMLS. However, these controlled vocabularies do not always cover the intended domain or have terms in the correct context of the desired NLP task. When existing sources cannot fit the needs of a designed task, annotation can provide the detail of the desired knowledge representation for creating semantic categories and lexicons. The existing coverage of PTSD terminology previous to the PTSDO development did not exist. While most of the existing terminology resources with PTSD coverage may be well

suited to the particular clinical setting for which they were developed, such application-specific representations limit the reuse of domain data. Most independent databases have overlapped data residing in dissimilar formats making integration costly in both monetary and time-based terms. Such barriers add to the challenge of retrieving information using common language in an understandable form [23]. PTSDO is hampered by the efficacy of readily accessible and insufficient annotated training data. This training data is necessary for other text mining strategies to generalize or make various predictions about new data [25]. Because robust validation is key to acceptance of a terminology, obtaining corpora for measuring performance of text mining pipeline and dictionary is vital. In this research, therefore, validation will play a vital role, and one that should help foster greater acceptance from the PTSD research community.

4.2.4 Corpus, Dictionary, and Pipeline Implementation

Annotated corpora or a collection of documents are vital to the evaluation of NER components including its referential terminology resource [326]. Terminology structures such as controlled vocabularies, taxonomies, and ontologies have played an important role in supporting NLP and NER systems as terminological knowledge, providing semantic constraints, and attaching meta-data [330]. In turn, these terminology structures can benefit themselves from concept recognition in text by identifying candidate terms missed that can be added to the controlled vocabulary or ontology [331].

Implementation of text mining pipelines in this project includes the execution of the application in order to annotate and extract relevant PTSD data related to variations in symptoms and therapeutic interventions. The pipelines enable the use of terminologies and biomedical ontologies in support of natural language processing (NLP) techniques. The output components generated by each include words, concepts, phrases, sentences and annotations of concepts. The dictionary component feeds into the named entity recognition module to annotate concepts. The text mining pipelines implemented for this paper include the NCBO Annotator [193], MetaMap [88], clinical Text Analysis and Knowledge Extraction System (cTAKES) [200], and a modified implementation cTAKES referred to as “SKATE.” SKATE stands for System for Knowledge Acquisition & Term Extraction and includes a regular expression module attached to the cTAKES pipeline. The clinical Text Analysis and Knowledge Extraction System (cTAKES), is released via open-source at http://www.ohnlp.org. It is a modular system that builds on existing

open-source technologies—the Unstructured Information Management Architecture framework and the OpenNLP NLP toolkit [200]. SKATE implements the pipelined components of cTAKES supported with additive learning components to provide both rule-based and dictionary based concept recognition. It also updates several processes of cTAKES to include a pre-processing regular expression stemming engine. NCBO Annotator is a web service provided by the National Center for Biomedical Ontology (NCBO) that annotates textual data with ontology terms from the UMLS and BioPortal ontologies [193]. The MetaMap is a highly configurable tool created to map biomedical text to the UMLS Metathesaurus. It parses input text into noun phrases for identification including their respective alternate spellings, abbreviations, synonyms, and inflections. MetaMap creates a candidate set of Metathesaurus terms and computes scores based upon strength of variant mapping to each candidate term [8, 190].

The terminology resources implemented to support text mining pipelines for this chapter include Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT) [97], National Cancer Institute (NCI) Thesaurus [100], PILOTS (Published International Literature on Traumatic Stress) Thesaurus [125], Unified Medical Language System (UMLS) [83], and the PTSDO. The UMLS is an aggregation of over 100 terminologies to provide characteristics of natural language, concept mapping, and semantic categorization [21]. SNOMED-CT is a clinical controlled vocabulary that aggregates medical terminology lists, hierarchical representation of concepts, and expressional definitional knowledge of terms [97-98]. The NCI Thesaurus provides resources and services to meet the NCI's needs for controlled terminology and to facilitate the standardization of terminology and information systems [100]. Many patients with cancer experience some form of post-traumatic stress therefore the NCI-T contains many concepts within PTSD [332]. The PILOTS Thesaurus is essentially a purpose-built controlled vocabulary used to index and retrieve literature [49] for a database of articles, books, posters, gray literature, pamphlets, white papers, and all materials of practical value for issues surrounding the disorder [126]. Lastly, PTSDO as described in Chapter 3 is implemented to support each of the pipelines. PTSDO is a terminology system developed from acquiring concepts and terms from existing resources both manually and implementing a semi-automatic discovery method. For this research and terminology development, a concept is used to describe meaning and a term is defined as the actual word [21, 68, 70]. For the remainder of the chapter, concept or term is used to refer to the tuple of namespace, identifier, definition, and synonyms of

the terminology resource. This chapter evaluates several hypotheses related to extracting concepts from various text processing pipelines as well as many dictionaries and analyzes the results of their respective combinations.

4.3 Description of Aim 2

Aim 2. Implement text mining pipelines supported with biomedical terminology resources for validation of PTSD concept coverage

4.3.1 Conceptual Model

Figure 4.2 describes the conceptual model of aim 2. A gold standard is created from the annotation of symptom and treatment concepts from two corpora. The first corpus is PubMed case reports and the second is psychotherapeutic transcripts for gold standard creation which is developed by manual annotation until a sufficient inter annotator agreement is achieved. Several text mining pipeline and terminology resource combinations are implemented for dictionary-based NER. Accuracy metrics of recall, precision, and F-measure are calculated and evaluated for significant difference between each combination.

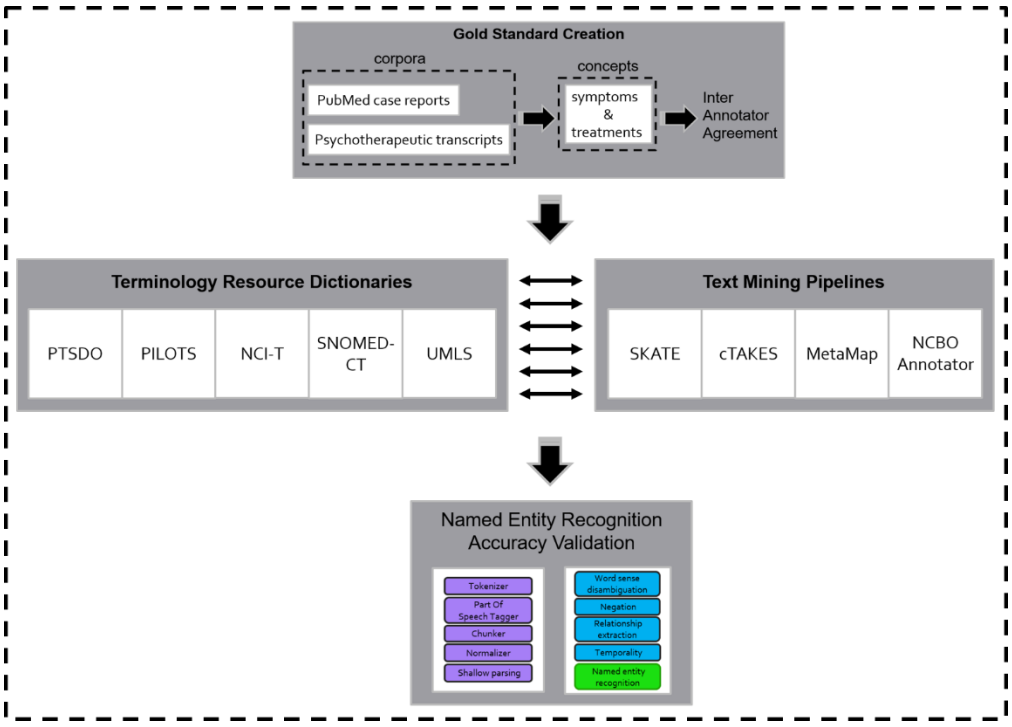


Figure 4.2. Conceptual model of aim 2.

4.3.2 Research Questions and Hypotheses

RQ3. *Do terminology resources containing PTSD concepts vary among text mining pipelines when processing biomedical corpora?*

Study 1: Automated annotation accuracy evaluation of symptom concepts in PubMed PTSD case reports with each dictionary and pipeline combination.

Study 2: Automated annotation accuracy evaluation of treatment concepts in PubMed PTSD case reports with each dictionary and pipeline combination.

Study 3: Automated annotation accuracy evaluation of symptom concepts in psychotherapeutic therapeutic session transcripts of PTSD patients with each dictionary and pipeline combination

Study 4: Automated annotation accuracy evaluation of treatment concepts in psychotherapeutic therapeutic session transcripts of PTSD patients with each dictionary and pipeline combination

Hypothesis 1 - Dictionary-based terminology resources will not perform equally on corpora with text processing pipelines.

Hypothesis 2 – The accuracy of dictionary coverage will significantly vary between text processing pipelines.

Hypothesis 3 – The accuracy of PTSDO will be significantly higher than coverage of other evaluated dictionaries with text processing pipelines.

4.4 Methods

This section provides an overview of the literature and clinical narrative corpora that was utilized in this research. It also presents the methodology and specific tasks for building two gold

standards in order to evaluate the PTSD terminology resources in combination with the text mining pipelines. Lastly, this section discusses the methods for pipeline implementation.

4.4.1 Description of Corpora

The first corpus utilized in this research is PTSD case reports downloaded from PubMed which is maintained by the National Library of Medicine (NLM) to provide access to MEDLINE. MEDLINE is a database of indexed biomedical journal articles to include biology, health care, selected life sciences journals articles, citations, and abstracts [333]. PubMed was searched with the search terms "PTSD case report" which returned 1202 citations with available abstracts. These case reports provided detailed descriptions of the patient's persistent symptoms and therapeutic interventions.

The second corpus consisted of counseling and psychotherapy transcripts, client narratives, and reference works that were downloaded from Alexander Street Press at: http://alexanderstreet.com/products/counseling-and-psychotherapy-transcripts-series. This collection contains more than 6,000 transcripts of client-therapist sessions, 40,000 pages of client narratives, and 25,000 pages of reference works across a broad array of symptoms and therapeutic strategies. To ensure accuracy, all therapists in provided transcripts must adhere to the American Psychological Association's (APA) Ethics Guidelines [334]. After filtering patients based upon diagnosis, 73 PTSD patients were identified with available transcripts. The number of therapy sessions for each patient ranged from a low of 1 to a high of 91 for a total of 642 available transcripts. To give an idea of the size of the files, the number of tokens or grouped sequence of characters within each document ranged from 287 to 14,909.

After accumulating all available PubMed PTSD case reports and psychotherapeutic transcripts, 10% was randomly selected and set aside for creating the gold standards. This testing set is not used in the development of the dictionary and pipeline training, therefore the results of this experiment will be indicative of overall performance. For the evaluation of identification of symptom and treatment terminology, 122 PubMed PTSD case report abstracts are used for testing. Psychotherapeutic therapy sessions are tested with 6 PTSD patients, which included 82 transcripts. This gold standard is focused toward PTSD entity extraction aimed to provide training and evaluation support for text mining applications.

4.4.2 Gold Standard Development

The creation of gold standards for terminology validation entails rigorous annotation of specified concepts. The methodology followed for the annotation process is developed by Boisen et al. and is specifically designed to optimize strategies for extracting entities. It follows standards established for NLP tasks in order to maximize results oriented annotations. The methodology discusses guideline creation, demands at least two annotators, determines acceptable levels of agreement and outlines adjudication [327]. Adjudication is the process of comparing the sets of annotated documents for a final approval consensus. Measurements are obtained based upon matches and non-matches and any discrepancies are resolved [327, 335].

The corpus consists of the set aside evaluation dataset for which a schema is created to provide clarification about domain requirements of the annotation task [335]. Annotation guidelines provide a sequence of steps in order to minimize errors and give explicit details on what should and should not be annotated. An annotation guideline is defined with the annotation classes of signs or symptoms, and therapeutic interventions. Three clinicians are trained to annotate the PubMed case reports and the psychotherapeutic therapy sessions according to the guideline. Each document is annotated by two annotators and the third annotator performed adjudication.

The two annotators and one adjudicator are all clinicians with at least some experience in medical informatics thus each has a good understanding for the needs of manually annotated corpora. The first annotator has a medical degree and experience in seeing patients that has been diagnosed with PTSD. The second annotator, though not an expert, has had prior experience with implementations of NLP systems. The adjudicator is a physician that has specialized training in PTSD with a focus in treating patients suffering from traumatic brain injury.

4.4.2.1 Annotation Guidelines

Consistency is important for quality annotated data and it starts with developed guidelines that adhere to design standards. The guidelines provide a sequence of steps in order to minimize errors [335]. There are specific references as to what text is included in an annotation set out in Bada et al. [135]. While a subset of annotation guidelines is attached in **Appendix D**, the general steps summarized for creation shown in **Table 4.2** aid in minimizing human error:

1	**Read the entire document.**	Read the document through in its entirety to get an understanding, marking no annotations.
2	**Identify entities.**	Read the document a second time, adding annotations for the mentions (including synonyms/pronouns) of the entities.
3	**Identify meta-data annotations.**	Check for modifiers, qualifications and co-references.
4	**Identify hierarchical relationships.**	Check if entities have identifiable IS-A relationships with other entities.
5	**Record any additional information.**	Record any comments, questions, uncertainties or ambiguities

Table 4.2. Summary of annotation guidelines [335]

The development of these annotation schema and guidelines is created in iterations based upon interviews and conversations with domain experts. The iterative process of the development is shown in **Figure 4.3**. At the most basic level, the guidelines state what should and should not be annotated. They guide annotation rules for overlapping terms, the breaking down of complex terms, conjunctions. Not every word requires a label but only those useful for the intended extraction task. Annotators were advised not collaborate when marking up documents and a given set should be completed by only one annotator.

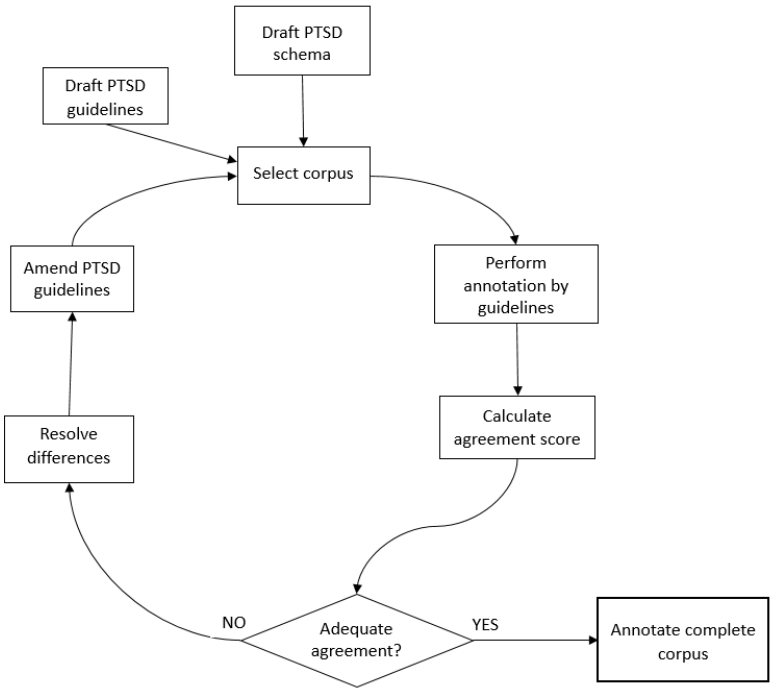


Figure 4.3. PTSD annotation schema and guideline iterative development.

Well-annotated documents increase the flexibility on the generality or specificity of queries [326]. Consistent annotation of text is difficult but it significantly increases its usefulness. The consistency increases understandability and enables reuse of the developed corpus for additional projects [136]. Especially in medical fields, at least some domain knowledge can provide better annotation. Without domain knowledge, more detail in the guidelines is required and training will be extensive and time-consuming [326]. The first task is the annotation of the symptomatology named entities in PubMed case report abstracts and psychotherapeutic transcripts. The second task consists of annotating therapeutic intervention named entities in PubMed case report abstracts and psychotherapeutic transcripts. The document count, range of concepts, and range of tokens is shown in **Table 4.3**.

	PubMed Case Reports	Psychotherapeutic Transcripts
Documents	122	82
Concepts	0 to 37	7 to 94
Tokens	126 to 2878	287 to 14,909

Table 4.3. Statistics for case reports and psychotherapeutic transcripts

Annotators were instructed to annotate the minimal amount of tokens necessary in order to accurately describe the identified PTSD concept. When questions arose, annotators were instructed to err on the side of annotating more which could be recognized and corrected by the adjudicator. Having direct access to existing terminology and knowledge sources kept these instances to a minimum. For this annotation task, Semantator (Version 1.0) annotation tool available at http://informatics.mayo.edu/CNTRO/index.php/Semantator. Semantator is compatible with Protege 4.1 (or newer). It provides an environment for browsing and querying the annotated data, as well as interactively refining annotation results if needed in the selected corpus of text. The guide is written for annotation tasks using the manual annotation mode. In this mode, a human expert curator can choose a document to be annotated and a domain terminology, highlight different pieces of information from the original text, and then mark which source the concepts belongs. The system will generate class instances according to curator's annotation and display different class instances that is color-coded.

4.4.2.2 Inter-annotator Agreement

Clinical concepts in text can be poorly defined, overlap with other important concepts, or not pertain to the manual annotation process due to contextual differences. The large number of true negatives prevalent in a named entity annotation task propagates calculating standard agreement measures such as kappa less meaningful if not impossible. In the building of these gold standards, the F-measure for inter-annotator agreement (IAA) is employed in order to avoid requiring the count of true negatives. The annotations of the first annotator are treated as the reference for evaluating differences of the second annotator in order to calculate the F-measure. This method allows the measure to approach the value of kappa when the conditional probability of one annotator agrees in a positive annotation given that the other annotator identified it as positive as well. Utilizing the F-score as discussed by Hripcsak et al. and shown in **Figure 4.4**, it calculates the harmonic mean of recall [336]. Therefore, the F-measure is utilized as the IAA in this research evaluation since it does not require a count of the number of negative cases [337].

$$F = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$$

Figure 4.4 F-score calculation

All agreements from the original annotators are accepted into a comprehensive set, then subsequently adjudicated on differences. The adjudicator is directed to not overrule either annotator to enforce their own single annotation nor do they create new annotations that have not been previously created [327]. Differences are resolved via reaching a consensus among the adjudicator and annotators. Clear and agreed upon mistakes within annotations were quickly corrected with collaboration among the group. A consensus was reached for the more complex annotations with group discussion, and a voted upon gold standard submission, with the adjudicator making the final decision. Several practice annotation sessions took place in order to familiarize annotators with the corpora, software, and guidelines. For example, from the text in (1), annotators identify the concepts in (2).

(1) **Cognitive behavioral therapy** for pediatric **insomnia** and **traumatic event** related **hyperarousal**.

(2) ***Cognitive behavioral therapy*** *= Therapeutic or Preventive Procedure class*
insomnia *= Sign or Symptom*
traumatic event *= Sign or Symptom*
hyperarousal *= Sign or Symptom*

A semantic type is a broad category described in the UMLS and assigned to each Metathesaurus concept. They are arranged in a hierarchy which is organized into two main categories, entity and event. Semantic types exist in differing levels of granularity or specificity [83-84]. Annotators identify concepts or phrases that are of the semantic types Therapeutic or Preventive Procedure and Sign or Symptom. They are directed to include any abbreviations or acronyms that represent the semantic type classes. Any questionable concept, phrase, abbreviation, or acronym is to be excluded. Included is any mention of negation in the proper annotation category. The signs and symptoms for the guidelines are only concerned with any mention of those related to PTSD, both asserted or negated. In an effort to capture symptoms for PTSD, the annotators were directed to not annotate any diagnoses of depression or major depressive disorder as a Sign or Symptom.

Annotators were not allowed to collaborate during each individual batch of annotations. However, the adjudicator was allowed to consult with each annotator for clarifications in the creation of the gold standard. The gold standard concepts are stored as JSON files. Examples from each corpus is shown in **Appendix E**. The default assertion for these tasks is annotations of

positive confirmations and related to the domain of PTSD. Unless otherwise stated, the annotators were directed to only annotate mentions that are present/confirmed and related to the disorder. For those cases where items are negated, the annotation must correspond to the negation category as requested. Spans must be one continuous string of text and may not be bridged when unmarkable text lies within. The span is defined as the minimal text span that captures the concept.

4.4.3 Pipeline and Dictionary Implementation

Four text mining pipelines are implemented: 1) NCBO Annotator; 2) MetaMap; 3) cTAKES; and 4) SKATE. Textual input for NCBO Annotator utilizes mgrep for annotation production, then converts into XML for pipeline evaluation [340]. MetaMap is installed locally for connecting to terminological resources using the MetaMap Data File Builder. The file builder requires terminologies to be formatted exactly as the locally installed UMLS which required a python script to convert each terminology source to compatible database tables [331]. cTAKES is an NLP system for information extraction from electronic medical record clinical free-text. It is built using the UIMA (Unstructured Information Management Architecture) framework and OpenNLP toolkit. The system is designed to identify named entities such as drugs, diseases/disorders, signs/symptoms, anatomical sites and procedures. cTAKES components are specifically trained for the clinical domain and include a sentence boundary detector, tokenizer, normalizer, part-of-speech tagger, phrasal chunker, dictionary lookup annotator, negation detector, and dependency parser [200]. Utilizing the UIMA framework, feature structure types are created in files for SKATE called ‘Type System Descriptor’ and ‘Analysis Engine Descriptor’ in order to map to XML elements. SKATE utilizes analysis engines of cTAKES but also implements Negex [201] for negation detection, and attaches a regular expression module for improving NER. The parameters for each system are set to the optimal levels as described in the analyses by Savova et al., Funk et al., Jonquet et al., and Pratt [193, 200, 331. 341].

Five terminology resources are formatted and converted into dictionaries for text mining pipeline support: 1) PTSDO; 2) PILOTS Thesaurus; 3) SNOMED-CT; 3) NCI Thesaurus; and 5) UMLS. PTSDO is modeled to include symptom and treatment vocabulary explicit to the domain of PTSD. The taxonomy contains 1,244 concepts and 1,338 terms. The current status of the PTSDO lexicon has been vetted by domain experts to correctly describe the concepts that exist

within the disorder. The PILOTS Thesaurus [124] is a purpose-built controlled vocabulary used to index and retrieve PTSD literature. The thesaurus contains a list of more than 1200 terms. SNOMED-CT [94] is a controlled vocabulary designed to aggregate concepts for clinical findings, symptoms, diagnoses, procedures, body structures, organisms, substances, pharmaceuticals, devices and specimens. SNOMED-CT contains 311,000 concepts and almost 800,000 terms. Consisting of 118,941 concepts, the NCI Thesaurus provides controlled terminology for clinical care, translational and basic research, and public information and administrative activities. The concepts include codes, terms, abbreviations, synonyms, definitions, links to outside sources, and additional research community supportive information [101]. The UMLS [83] provides a major source of clinical domain concepts and terms in ontologies and controlled terminologies retrievable from a relational database. The current count of the concepts in the UMLS is 3,250,228 concepts, and 12,997,673 terms. The database requires a great amount of scripting into to convert into XML-based format for text mining support. Each terminology resource is implemented to provide dictionary support to each pipeline for a total of 20 combinations.

4.4.4 Evaluation pipeline

An evaluation pipeline for each system is constructed in Python. The calculations and scoring algorithm is shown in **Appendix H**. The accuracy evaluation consists of recall **Figure 4.5a**, precision **Figure 4.5b**, and F-measure **Figure 4.5c**. The gold standard concepts are stored as JSON files of which an excerpt is shown in **Appendix E**.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Figure 4.5a. Recall formula

$$\text{Precision} = \frac{TP}{TP + FP}$$

Figure 4.5b. Precision formula

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

Figure 4.5c. F-measure formula

The definitions of what constitutes an entity identification boundary is shown in **Table 4.4**. The correct annotation is illustrated in the table followed by an example NER output. Precision, recall and F-score are computed for exact and partial matches of identified named entities. An exact match consists of each token in the gold standard annotation. Partial matches are important for capturing information because of multiple annotations per span of text, nested spans, and overlapping spans of textual data. The exact match and left boundary match requires either all

tokens or a complete left-side token to recognized. The same is true for the exact and right boundary matches but for the right-side token. For example, system identification of 'traumatic' in the gold standard annotation of 'traumatic hallucination symptom' is counted as a left boundary match and included into all matches. A non-boundary match includes any token within the right and left boundary token. The boundary matches consist of the beginning offset of the named entity found by the pipeline-dictionary combination that matches the beginning offset of the manual annotation. Each left, right, and non-boundary entity identification is not subsumed by the exact match. The all matches scoring metric subsumes each matching category and is counted if any of the matching rules are met, meaning the gold stand annotation is recognized in some token form.

	Gold standard annotation	NER output
Exact matches	"traumatic hallucination symptom"	"traumatic hallucination symptom"
Exact matches and left boundary matches	"traumatic hallucination symptom"	"traumatic"
Exact matches and right boundary matches	"traumatic hallucination symptom"	"symptom"
Exact matches and non-boundary matches	"traumatic hallucination symptom"	"hallucination"
All matches*	"traumatic hallucination symptom"	Meets 1 or more of the above matches.

*mutually exclusive

Table 4.4. Entity identification boundary definition

In order to determine statistical difference in text mining pipeline implementations across each of the dictionary inputs, Kruskal-Wallis [342] determines whether there are significant differences in the F-measures of all boundary matches [331]. The test is a rank-based nonparametric test and chosen because the collected data has similar shape but does not follow a normal distribution. In these experiments, it is appropriate because the goal is to determine if the distribution of scores are from a particular experimental condition. The variable condition in this statistical analyses is each of the dictionary inputs. All assumptions of Kruskal-Wallis are met and analysis is conducted using the Stata Data Analysis and Statistical Software Package. Significance is determined at a 95% level, $\alpha = 0.05$. For each document in both corpora inputted,

the mean and variance is computed across dictionary-pipeline combination outputs. These calculations are computed at the document-level using the F1 measure for input into Stata, then at the corpus-level using a micro-average. The null hypothesis in each pairwise comparison is that the probability of observing a random value in the first group that is larger than a random value in the second group equals one-half. Upon rejection of the null hypothesis of no statistical difference between each combination grouping, the Dunn's pairwise comparison with a Bonferroni correction is implemented. With the Bonferroni multiple comparison for 4 pipelines across 5 dictionary inputs, the new significance level is $\alpha = 0.00026$. As in the rank-sum test, if the data are assumed to be continuous and the distributions are assumed to be identical except for a shift in centrality, Dunn's test may be understood as a test for median difference. These multiple comparisons are produced using the dunntest command in Stata following its built-in omnibus kwallis command [343].

In order to reduce false positives (FPs), the symptom concept identification is confined to the semantic type categories in **Table 4.5a**. A large number of concepts are of the types 'Sign or Symptom' and 'Findings.' A number of concepts is missed in identification but limiting recognition to these types increases overall NER performance by greatly reducing the false positives (FPs) in comparison to missed true positives (TPs). Similar to symptom semantic type confinement, **Table 4.5b** displays the treatment semantic types the concept recognition systems are permitted to identify.

Sign or Symptom	T184	sosy
Mental Process	T041	menp
Finding	T033	fndg
Individual Behavior	T055	inbe
Mental or Behavioral Dysfunction	T048	mobd
Social Behavior	T054	socb
Injury or Poisoning	T037	inpo

Table 4.5a. Semantic types utilized for symptom concept recognition

Therapeutic or Preventive Procedure	T061	topp
Mental Process	T041	menp
Finding	T033	fndg
Educational Activity	T065	edac
Health Care Activity	T058	hlca
Idea or Concept	T078	idcn
Intellectual Product	T170	inpr
Pharmacologic Substance	T121	phsu
Individual Behavior	T055	inbe

Table 4.5b. Semantic types utilized for treatment concept recognition

4.5 Results

4.5.1 Gold Standard Creation

This section provides details of the gold standard creation and results of the completed annotation. The corpora for annotation of symptom and treatment concepts consist of 122 PubMed case reports and 82 psychotherapeutic transcripts. The results presented are post-training of four iterations of practice sessions with annotators. Annotations are based upon exact matches, left-boundary, right-boundary, partial, and cumulative identification corresponding to the entity extraction implementation with text mining pipelines. An example of a treatment annotation is shown below in **Figure 4.6**.

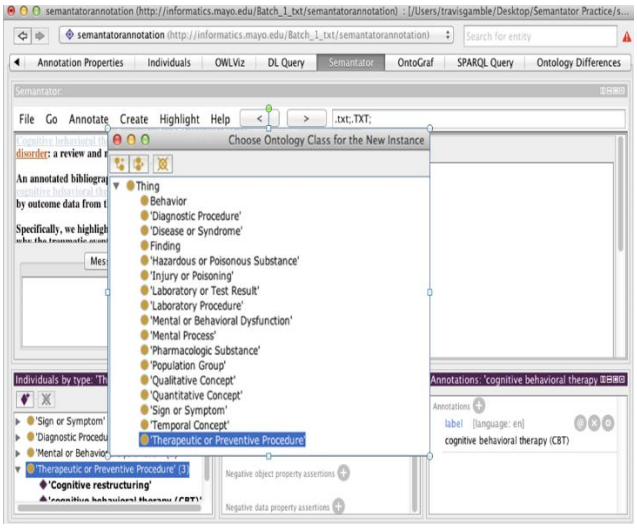


Figure 4.6. Sample annotation from Semantator

Gold standard inter-annotator agreement (IAA) is excellent for both symptoms and treatments annotations of PubMed case reports and psychotherapeutic transcripts. The IAA results for the annotations are presented in **Table 4.6**. After adjudication, the exact matches of symptom concepts from PubMed abstracts (F=0.80) improves for all boundary matches (F=0.89) versus the adjudicated set. The majority of the improvements derives from exact and right-side boundary matches (F=0.86). For treatment concepts in PubMed abstracts, there is less improvement from exact matches (F=0.89) to all boundary matches (F=0.91) versus the adjudicated set. In contrast, exact matches of symptom concepts from psychotherapeutic transcripts (F=0.68) greatly improves for all boundary matches (F=0.82). Treatment concepts in the transcripts slightly improves from all matches (F=0.81) to all boundary matches (F=0.84).

The exact matches of symptom concepts from PubMed abstracts (F=0.80) is much higher than the symptom concepts from psychotherapeutic transcripts (F=0.68) while minimal differences are recognized in treatment concepts.

	PubMed abstracts		Psychotherapeutic trans.	
	Sx	Tx	Sx	Tx
Exact matches	0.80	0.89	0.68	0.81
Exact matches and left boundary matches	0.82	0.90	0.76	0.84
Exact matches and right boundary matches	0.86	0.89	0.77	0.83
Exact matches and non-boundary matches	0.82	0.90	0.72	0.83
All matches	0.89	0.91	0.82	0.84

Table 4.6. Gold standard results inter-rater agreement

The adjudicator examined the annotations of both annotators, resolving differences and determined the organization of the final gold standard. The total count of symptom concepts in PubMed abstracts is determined to be 987 and 314 for treatment concepts. The average count for each abstract is 6 concepts. The range of PubMed concept count is 0 to 37. The total count of symptom concepts in psychotherapeutic transcripts is determined to be 1,648 and 723 for treatment concepts. The average number of concepts for each transcript is 13 and the range is 7 to 94 concepts.

4.5.2 Dictionary and Pipeline Evaluation

The results from the corpora named entity recognition (NER) with the collective dictionaries and pipelines are presented in **Figures 4.7** through **4.26**. The figures display each boundary's concept recognition metrics of recall, precision, and F-measure. A summary across each pipeline and dictionary, including the count of true-positives, false-positives, and false-negatives are attached in **Appendix J**. The scale of NER performance will reference F-measures from 0.00 to 0.35 as being poor, from 0.35 to 0.55 as moderate, between 0.55 to 0.80 as good, from 0.80 to 0.95 as excellent, and any F-measure above 0.95 as state of the art.

4.5.2.1 Evaluation 1

Dictionary: PTSDO Corpus: PubMed abstracts Target: Symptom

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.70	0.93	0.80	0.83	0.94	0.88	0.76	0.93	0.84	0.75	0.92	0.83
Exact and left boundary matches	0.77	0.93	0.84	0.89	0.95	0.92	0.81	0.93	0.87	0.81	0.93	0.86
Exact and right boundary matches	0.76	0.93	0.84	0.87	0.95	0.90	0.81	0.93	0.87	0.81	0.93	0.87
Exact and non-boundary matches	0.72	0.93	0.81	0.85	0.95	0.90	0.78	0.93	0.85	0.78	0.93	0.85
All matches	0.81	0.94	0.87	0.92	0.95	0.94	0.84	0.94	0.89	0.84	0.93	0.88

Figure 4.7. Results from PTSDO and PubMed symptom analysis.

Shown in **Figure 4.7**, the F-measure with PTSDO dictionary and cTAKES pipeline combination is good for exact boundary matches (F = 0.80) and is excellent for all exact and any boundary match (F = 0.87). Both exact, left, and right boundary matches (F = 0.84) improved contributing to the success of the all matches accuracy. Overall, PTSDO-cTAKES recall is respectable for exact matches (R = 0.70) but greatly improved for all matches (R = 0.81). The highest non-exact match recall for this combination is with the left boundary (R = 0.77). PTSDO-cTAKES precision is great with this combination for exact boundary matches (P = 0.93). The dictionary-pipeline grouping only slightly improves for any match (P = 0.94). PTSDO and SKATE pipeline F-measure is excellent for exact boundary matches (F = 0.88). The highest accuracy of all pipelines supported with this terminology resource is this measure of all matches (F = 0.94) which is greatly supported with the left boundary (F = 0.92). Recall and precision are excellent for exact matches (R = 0.83); (P = 0.94), but tremendously improves for all matches (R = 0.92); (P = 0.95). The highest non-exact match recall is with the left boundary (R = 0.89). The F-measure with PTSDO dictionary and NCBO Annotator pipeline combination is excellent for exact boundary matches (F = 0.84) and improves for exact and any boundary match (F = 0.89). Both left and right boundary matches (F = 0.87) contributes to the improvement of accuracy**.** Recall is good for exact matches (R = 0.76) and improves for all matches (R = 0.84). Precision is very high with this combination and close to equal for boundary matches (P = 0.93) but slightly improved for any match (P = 0.94). The F-measure with PTSDO dictionary and MetaMap pipeline combination is excellent for exact boundary matches (F = 0.83) and is superior for all exact and any boundary match (F = 0.88). Recall is great for exact matches (R =

0.75) and improves for all matches (R = 0.84). The highest non-exact match recall is with both the left and right boundary (R = 0.81). Precision is extremely high with this combination for exact boundary matches (P = 0.92). The metrics only slightly improves for any match (P = 0.93).

4.5.2.2 Evaluation 2

Corpus: PubMed abstracts Target: Symptom Dictionary: PILOTS

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.72	0.90	0.80	0.74	0.90	0.82	0.72	0.89	0.80	0.76	0.87	0.82
Exact and left boundary matches	0.79	0.90	0.84	0.82	0.91	0.87	0.78	0.89	0.83	0.85	0.89	0.87
Exact and right boundary matches	0.77	0.90	0.83	0.80	0.91	0.85	0.76	0.89	0.82	0.82	0.88	0.85
Exact and non-boundary matches	0.74	0.90	0.81	0.76	0.91	0.83	0.74	0.89	0.81	0.78	0.88	0.83
All matches	0.80	0.91	0.85	0.84	0.91	0.88	0.79	0.89	0.84	0.87	0.89	0.88

Figure 4.8. Results from PILOTS and PubMed symptom analysis.

Figure 4.8 displays PILOTS dictionary and cTAKES pipeline combination with exact boundary matches (F = 0.80) but improves for all exact and any boundary match (F = 0.85). Recall is respectable for exact matches (R = 0.72) but greatly improves for all matches (R = 0.80). Precision is excellent with this combination for both exact (P = 0.90) and all matches (P = 0.91). The F-measure with PILOTS dictionary and SKATE pipeline combination is great for exact boundary matches (F = 0.82) and is outstanding for all exact and any boundary match (F = 0.88). Much of the accuracy of this measure improves with the left boundary match (F = 0.87). Recall is good for exact matches (R = 0.74) and greatly improves for all matches (R = 0.84). The majority of the improvement is with the left boundary match (R = 0.82). Precision with this combination for exact boundary matches (P = 0.90) is excellent. Collectively, left, right, and all boundary matches (P = 0.91) slightly improves. The F-measure with PILOTS dictionary and SKATE pipeline combination is great for exact boundary matches (F = 0.80) and is outstanding for all exact and any boundary match (F = 0.84). Overall, recall is decent for exact matches (R = 0.72) and slightly improves for all matches (R = 0.79). Precision with this combination remains consistent for exact, left, right, non-boundary and all matches (P = 0.89) which is excellent. In **Figure 4.8**, the F-measure with PILOTS dictionary and SKATE pipeline combination is great for exact boundary matches (F = 0.82) and is outstanding for all exact and any boundary match (F = 0.88). Much of the accuracy of this measure is improved with the left boundary match (F = 0.87).

Overall, recall is good for exact matches (R = 0.76) and greatly improves for all matches (R = 0.87). Precision with this combination for exact boundary matches (P = 0.87) is excellent and slightly improves for all matches (R = 0.89).

4.5.2.3 Evaluation 3

Corpus: PubMed abstracts Target: Symptom Dictionary: SNOMED-CT

	cTAKES			SKATE			NCBO			MetaMap		
	R	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
Exact matches	0.60	0.78	0.68	0.64	0.82	0.72	0.61	0.83	0.70	0.70	0.79	0.74
Exact and left boundary matches	0.70	0.80	0.75	0.77	0.84	0.81	0.76	0.86	0.81	0.83	0.82	0.82
Exact and right boundary matches	0.60	0.78	0.68	0.64	0.82	0.72	0.68	0.84	0.75	0.76	0.80	0.78
Exact and non-boundary matches	0.62	0.78	0.69	0.66	0.82	0.73	0.65	0.84	0.73	0.72	0.79	0.78
All matches	0.71	0.80	0.75	0.78	0.84	0.81	0.78	0.86	0.82	0.86	0.82	0.84

Figure 4.9. Results from SNOMED-CT and PubMed symptom analysis.

The F-measure with SNOMED-CT dictionary and cTAKES pipeline combination shown in **Figure 4.9** is good for exact boundary matches (F = 0.68). This combination improved with the addition of the left boundary match (F = 0.75) and made up the entire F-measure of all matches (F = 0.75). Generally, recall is decent for exact matches (R = 0.60) but greatly improves for all matches (R = 0.71). Precision is great with this combination for exact boundary matches (P = 0.78) and slightly improves for any match (P = 0.80). The F-measure with SNOMED-CT dictionary and SKATE pipeline combination is good for exact boundary matches (F = 0.72) and improves for all exact and any boundary match (F = 0.81). The majority of the accuracy of this measure improves with the left boundary match (F = 0.81). In **Figure 4.9**, recall was respectable for exact matches (R = 0.64) but tremendously improved for all matches (R = 0.78). The vast majority of recall for SNOMED-CT-SKATE is improved with the left boundary match (R = 0.77). Precision is great with this combination for exact boundary matches (P = 0.82). It slightly improves for both left boundary (P = 0.77) and any match (P = 0.78). SNOMED-CT-NCBO Annotator pipeline combination F-measure is good for exact boundary matches (F = 0.70) but tremendously improves for all exact and any boundary match (F = 0.82). Recall is acceptable for exact matches (R = 0.61) but tremendously improved for all matches (R = 0.78). The highest non-exact match recall was with the left boundary (R = 0.76). Precision is great with this

combination for exact boundary matches (P = 0.83) and for any match (P = 0.86). The F-measure with SNOMED-CT and MetaMap pipeline combination is good for exact boundary matches (F = 0.74) and greatly improves for all exact and any boundary match (F = 0.84). Recall is respectable for exact matches (R = 0.70) but tremendously improved for all matches (R = 0.86). Precision is good with this combination for exact boundary matches (P = 0.79).

4.5.2.4 Evaluation 4

Corpus: PubMed abstracts Target: Symptom Dictionary: NCI Thesaurus

	cTAKES			**SKATE**			**NCBO**			**MetaMap**		
	R	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
Exact matches	0.56	0.75	0.64	0.60	0.78	0.68	0.57	0.74	0.65	0.64	0.75	0.69
Exact and left boundary matches	0.61	0.76	0.68	0.67	0.80	0.73	0.69	0.78	0.73	0.70	0.76	0.73
Exact and right boundary matches	0.59	0.75	0.66	0.62	0.79	0.69	0.69	0.78	0.73	0.70	0.77	0.73
Exact and non-boundary matches	0.59	0.75	0.66	0.62	0.78	0.69	0.60	0.75	0.67	0.65	0.75	0.70
All matches	0.63	0.77	0.69	0.67	0.80	0.73	0.73	0.79	0.76	0.76	0.78	0.77

Figure 4.10. Results from NCI Thesaurus and PubMed symptom analysis.

The F-measure with NCI Thesaurus and cTAKES pipeline combination is moderate for exact boundary matches (F = 0.64) and only slightly improves for any boundary match (F = 0.69) as shown in **Figure 4.10**. Overall, recall is moderate for exact matches (R = 0.56) and slightly improves for all matches (R = 0.63). Precision is good with this combination for exact boundary matches (P = 0.75) and only slightly improves for any match (P = 0.77). NCI Thesaurus dictionary and SKATE pipeline F-measure combination is good for exact boundary matches (F = 0.68). It slightly improved for both left boundary (F = 0.73) and any match (F = 0.73). Recall is respectable for exact matches (R = 0.60). The accuracy of this measure is improved with the left boundary match (R = 0.67) equaling all matches (R = 0.67). Precision is great with this combination for exact boundary matches (P = 0.78). It only slightly improved with the left boundary match (P = 0.80) equaling all matches (P = 0.80). The F-measure with NCI Thesaurus and NCBO Annotator pipeline combination is good for exact boundary matches (F = 0.65) and improves greatly for any boundary match (F = 0.76)**.** Both left and right boundary matches (F = 0.73) improves accuracy. Recall is poor for exact matches (R = 0.57) but tremendously improves for all matches (R = 0.73). Precision is good with this combination for exact boundary matches

(P = 0.74). It slightly improved for any match (P = 0.79). NCI Thesaurus dictionary and MetaMap pipeline combination F-measure is decent for exact boundary matches (F = 0.69) but improves to a respectable (F = 0.77) for all exact and any boundary match. Overall, recall is acceptable for exact matches (R = 0.64) but greatly improves for all matches (R = 0.76). Precision is great with this combination for exact boundary matches (P = 0.75).

4.5.2.5 Evaluation 5

Corpus: PubMed abstracts Target: Symptom Dictionary: UMLS

	cTAKES			**SKATE**			**NCBO**			**MetaMap**		
	R	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
Exact matches	0.69	0.73	0.71	0.71	0.76	0.74	0.78	0.74	0.76	0.78	0.73	0.76
Exact and left boundary matches	0.73	0.74	0.74	0.84	0.79	0.81	0.83	0.76	0.79	0.84	0.75	0.79
Exact and right boundary matches	0.70	0.73	0.71	0.74	0.77	0.76	0.83	0.76	0.79	0.84	0.75	0.79
Exact and non-boundary matches	0.70	0.73	0.71	0.72	0.76	0.74	0.80	0.75	0.77	0.80	0.74	0.77
All matches	0.73	0.74	0.74	0.85	0.79	0.82	0.86	0.76	0.81	0.87	0.75	0.81

Figure 4.11. Results from UMLS and PubMed symptom analysis.

In **Figure 4.11**, the F-measure with UMLS-cTAKES combination is good for exact boundary matches (F = 0.71). This combination improves with the addition of the left boundary match (F = 0.74) and made up the entire F-measure of all matches (F = 0.74). Recall is good for exact matches (R = 0.69) and improves for all matches (R = 0.74). Precision is great with this combination for exact boundary matches (P = 0.73). The F-measure with UMLS dictionary and SKATE pipeline combination is excellent for exact boundary matches (F = 0.74) and slightly improves for any boundary match (F = 0.82). Overall, recall was respectable for exact matches (R = 0.71) but tremendously improves for all matches (R = 0.85). The vast majority of the accuracy of this measure is improved with the left boundary match (R = 0.84). Precision is good with this combination for exact boundary matches (P = 0.76). It only slightly improves for any match (P = 0.79) with the highest match on the left boundary (P = 0.79). UMLS-NCBO Annotator combination is good for exact boundary matches (F = 0.76) and respectively improves for all exact and any boundary match (F = 0.81). Shown in **Figure 4.11**, recall is good for exact matches (R = 0.78) but greatly improved for all matches (R = 0.86). Precision is great with this combination for exact boundary matches (P = 0.74) and only slightly improves for any match (P = 0.76). The F-measure with UMLS dictionary and MetaMap pipeline combination is good for

exact boundary matches (F = 0.76) and is excellent for all exact and any boundary match (F = 0.81). UMLS-MetaMap recall is great for exact matches (R = 0.78) but greatly improves for all matches (R = 0.87). Precision is good with this combination for exact boundary matches (P = 0.73).

4.5.2.6 Evaluation 6

Corpus: PubMed abstracts Target: Treatment Dictionary: PTSDO

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.53	0.75	0.62	0.57	0.75	0.65	0.55	0.71	0.62	0.57	0.75	0.65
Exact and left boundary matches	0.64	0.78	0.70	0.71	0.79	0.75	0.67	0.74	0.71	0.68	0.79	0.73
Exact and right boundary matches	0.66	0.79	0.72	0.77	0.80	0.78	0.69	0.75	0.72	0.71	0.79	0.75
Exact and non-boundary matches	0.55	0.75	0.63	0.61	0.76	0.68	0.57	0.71	0.63	0.58	0.76	0.66
All matches	0.71	0.80	0.75	0.83	0.81	0.82	0.73	0.76	0.75	0.74	0.80	0.77

Figure 4.12. Results from PTSDO and PubMed treatment analysis.

Displayed in **Figure 4.12**, the F-measure with PTSDO dictionary and cTAKES pipeline combination is poor for exact boundary matches (F = 0.62) but greatly improves for all exact and any boundary match (F = 0.75). Recall is poor for exact matches (R = 0.53) but greatly improves for all matches (R = 0.71). The majority of the additional detection is recognized by the left boundary (R = 0.64) and the right boundary matches (R = 0.66). Precision is good with this combination for exact boundary matches (P = 0.75) but improves for any match (P = 0.80). The F-measure with PTSDO dictionary and SKATE pipeline combination is poor for exact boundary matches (F = 0.65) but tremendously improves for all exact and any boundary match (F = 0.82). The majority of the additional detection is recognized by the left boundary (F = 0.75) and the right boundary matches (F = 0.78). Recall is poor for exact matches (R = 0.57) but greatly improves for all matches (R = 0.83). The highest non-exact match recall was with the right boundary (R = 0.77). Precision is good with this combination for exact boundary matches (P = 0.75) and improves for any match (P = 0.82). The F-measure with PTSDO dictionary and NCBO Annotator pipeline combination is not good for exact boundary matches (F = 0.62) and improves for all exact and any boundary match (F = 0.75). The majority of the additional detection is recognized by the left boundary (F = 0.71) and the right boundary matches (F = 0.72). Recall is poor for exact matches (R = 0.55) but greatly improves for all matches (R = 0.73). Precision is

good with this combination for exact and boundary matches (P = 0.71) and improves for any match (P = 0.76). PTSDO-MetaMap is not good for exact boundary matches (F = 0.65) but improves for all exact and any boundary match (F = 0.77). The majority of the additional detection in **Figure 4.12** is recognized by the left boundary (F = 0.73) and the right boundary matches (F = 0.75). Recall is poor for exact matches (R = 0.57) but greatly improves for all matches (R = 0.74). The majority of the additional detection is recognized by the left boundary (R = 0.68) and the right boundary matches (R = 0.71). Precision is good with this combination for exact and boundary matches (P = 0.75) and improves for any match (P = 0.80).

4.5.2.7 Evaluation 7

Corpus: PubMed abstracts Target: Treatment Dictionary: PILOTS

	cTAKES			**SKATE**			**NCBO**			**MetaMap**		
	R	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
Exact matches	0.27	0.57	0.36	0.28	0.57	0.37	0.27	0.52	0.36	0.27	0.50	0.35
Exact and left boundary matches	0.38	0.65	0.48	0.43	0.67	0.52	0.41	0.62	0.49	0.42	0.60	0.49
Exact and right boundary matches	0.37	0.65	0.47	0.40	0.65	0.50	0.38	0.60	0.47	0.40	0.59	0.47
Exact and non-boundary matches	0.30	0.60	0.40	0.32	0.60	0.42	0.31	0.55	0.40	0.31	0.53	0.39
All matches	0.42	0.68	0.52	0.48	0.69	0.57	0.45	0.64	0.52	0.46	0.63	0.53

Figure 4.13. Results from PILOTS and PubMed treatment analysis.

Figure 4.13 shows that PILOTS dictionary and cTAKES pipeline combination F-measure is very poor for exact boundary matches (F = 0.36) and improves for all exact and any boundary match (F = 0.52). The majority of the additional detection is recognized by the left boundary (F = 0.48) and the right boundary matches (F = 0.48). Recall is poor for exact matches (R = 0.27) and slightly improves for all exact and any boundary match (R = 0.42). The majority of the additional detection is recognized by the left boundary (R = 0.38) and the right boundary matches (R = 0.37). Precision is poor with this combination for exact boundary matches (P = 0.57) but improves for any match (P = 0.68). The F-measure with PILOTS dictionary and SKATE pipeline combination is very poor for exact boundary matches (F = 0.37) and improves for all exact and any boundary match (F = 0.57). The majority of the additional detection is recognized by the left boundary (F = 0.52) and the right boundary matches (F = 0.50). Recall is poor for exact matches (R = 0.28) and improves for all exact and any boundary match (R = 0.48). Precision is poor with this combination for exact boundary matches (P = 0.57) but improves for any match (P = 0.69).

The F-measure with PILOTS dictionary and NCBO Annotator pipeline combination is very poor for exact boundary matches (F = 0.36) and improves for all exact and any boundary match (F = 0.52). The majority of the additional detection is recognized by the left boundary (F = 0.49) and the right boundary matches (F = 0.47). Recall is poor for exact matches (R = 0.27) and improves for all exact and any boundary match (R = 0.45). Shown in **Figure 4.13**, he majority of the additional detection is recognized by the left boundary (R = 0.41) and the right boundary matches (R = 0.38). Precision is poor with this combination for exact boundary matches (P = 0.52) but improves for any match (P = 0.64). PILOTS-MetaMap is very poor for exact boundary matches (F = 0.35) and improves for all exact and any boundary match (F = 0.53). Recall is poor for exact matches (R = 0.27) and improves for all exact and any boundary match (R = 0.46). Precision is poor with this combination for exact boundary matches (P = 0.50) but improves for any match (P = 0.63).

4.5.2.8 Evaluation 8

Corpus: PubMed abstracts Target: Treatment Dictionary: SNOMED-CT

	cTAKES			**SKATE**			**NCBO**			**MetaMap**		
	R	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
Exact matches	0.27	0.39	0.32	0.29	0.39	0.33	0.27	0.38	0.32	0.34	0.40	0.37
Exact and left boundary matches	0.52	0.55	0.53	0.56	0.55	0.55	0.51	0.53	0.52	0.65	0.56	0.60
Exact and right boundary matches	0.54	0.56	0.55	0.56	0.55	0.55	0.48	0.52	0.50	0.66	0.56	0.61
Exact and non-boundary matches	0.46	0.52	0.49	0.49	0.51	0.50	0.43	0.49	0.46	0.46	0.47	0.46
All matches	0.67	0.61	0.64	0.70	0.60	0.64	0.66	0.60	0.63	0.73	0.59	0.65

Figure 4.14. Results from SNOMED-CT and PubMed treatment analysis.

The F-measure with SNOMED-CT dictionary and cTAKES pipeline combination as shown in **Figure 4.14** is very poor for exact boundary matches (F = 0.32) but greatly improves for all exact and any boundary match (F = 0.64). The majority of the additional detection is recognized by the left boundary (F = 0.53) and the right boundary matches (F = 0.55). Recall is very poor for exact matches (R = 0.27) but improves for all exact and any boundary match (R = 0.67). The majority of the additional detection is recognized by the left boundary (R = 0.52) and the right boundary matches (R = 0.54). Precision is also very poor with this combination for exact boundary matches (P = 0.39) but greatly improves for any match (P = 0.61). The F-measure with SNOMED-CT dictionary and SKATE pipeline combination is very poor for exact boundary

matches (F = 0.33) but greatly improves for all exact and any boundary match (F = 0.64). Shown in **Figure 4.14**, recall is very poor for exact matches (R = 0.29) but tremendously improves for all matches (R = 0.70). Precision is very poor with this combination for exact boundary matches (P = 0.39) but improves for any match (P = 0.60). SNOMED-CT dictionary and NCBO Annotator pipeline combination F-measure is very poor for exact boundary matches (F = 0.32) but improves for all exact and any boundary match (F = 0.63). Overall, recall is very poor for exact matches (R = 0.27) but improves for all matches (R = 0.66). Left boundary (R = 0.51), right boundary (R = 0.48), and non-boundary (R = 0.43) each contributed to the any match measure (R = 0.66). Precision is very poor with this combination for exact and boundary matches (P = 0.38) but greatly improves for any match (P = 0.60). The F-measure with SNOMED-CT dictionary and MetaMap pipeline combination is very poor for exact boundary matches (F = 0.37) but greatly improves for all exact and any boundary match (F = 0.65). The majority of the additional detection is recognized by the left boundary (F = 0.60) and the right boundary matches (F = 0.61). Recall is very poor for exact matches (R = 0.34) but greatly improves for all matches (R = 0.73). Left boundary (R = 0.65), right boundary (R = 0.66), and non-boundary (R = 0.46) each contributed to the any match measure (R = 0.73). Precision is poor with this combination for exact and boundary matches (P = 0.38) but greatly improves for any match (P = 0.59).

4.5.2.9 Evaluation 9

Corpus: PubMed abstracts Target: Treatment Dictionary: NCI Thesaurus

	cTAKES			**SKATE**			**NCBO**			**MetaMap**		
	R	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**	**R**	**P**	**F**
Exact matches	0.27	0.53	0.36	0.29	0.49	0.36	0.24	0.41	0.31	0.30	0.43	0.36
Exact and left boundary matches	0.58	0.71	0.64	0.60	0.67	0.64	0.52	0.60	0.56	0.65	0.62	0.64
Exact and right boundary matches	0.68	0.74	0.71	0.70	0.70	0.70	0.63	0.65	0.64	0.71	0.64	0.67
Exact and non-boundary matches	0.48	0.66	0.55	0.48	0.62	0.54	0.43	0.56	0.49	0.54	0.58	0.56
All matches	0.73	0.75	0.74	0.76	0.72	0.74	0.65	0.65	0.65	0.77	0.66	0.71

Figure 4.15. Results from NCI Thesaurus and PubMed treatment analysis.

Displayed in **Figure 4.15**, the F-measure with SNOMED-CT dictionary and cTAKES pipeline combination is very poor for exact boundary matches (F = 0.36) but tremendously improves for all exact and any boundary match (F = 0.74). Left boundary (F = 0.64), right boundary (F = 0.71), and non-boundary (F = 0.55) each contributed to the any match measure (F

= 0.74). Recall is very poor for exact matches (R = 0.27) but greatly improves for all matches (R = 0.73). Left boundary (R = 0.58), right boundary (R = 0.68), and non-boundary (R = 0.48) each contributed to the any match measure (R = 0.73). Precision is poor with this combination for exact boundary matches (P = 0.53) but greatly improves for any match (P = 0.75). SNOMED-CT dictionary and SKATE pipeline combination F-measure is very poor for exact boundary matches (F = 0.36) but tremendously improves for all exact and any boundary match (F = 0.74). Recall is very poor for exact matches (R = 0.29) but tremendously improves for all matches (R = 0.76). Precision is very poor with this combination for exact boundary matches (P = 0.49) but improves for any match (P = 0.72). NCI Thesaurus and the NCBO Annotator pipeline F-measure is very poor for exact boundary matches (F = 0.31) but improves for all exact and any boundary match (F = 0.65). Left boundary (F = 0.56), right boundary (F = 0.64), and non-boundary (F = 0.49) matches did show F-measure improvement contributing to increasing all matches accuracy. Overall, recall is very poor for exact matches (R = 0.24) but improves for all matches (R = 0.65). Left boundary (R = 0.52), right boundary (R = 0.63), and non-boundary (R = 0.43) each contributed to the any match measure. Precision is poor with this combination for exact and boundary matches (P = 0.41) but improves for all matches (P = 0.65). Left boundary (P = 0.60), right boundary (P = 0.65), and non-boundary (R = 0.56) each contributed to the any match measure. In **Figure 4.15**, the F-measure with NCI Thesaurus dictionary and MetaMap pipeline combination is very poor for exact boundary matches (F = 0.36) but improves for all exact and any boundary match (F = 0.71). Left boundary (F = 0.64), right boundary (F = 0.67), and non-boundary (F = 0.56) matches did show F-measure improvement contributing to increasing all matches accuracy. Recall is very poor for exact matches (R = 0.30) but greatly improves for all matches (R = 0.77). Left boundary (R = 0.65), right boundary (R = 0.71), and non-boundary (R = 0.54) each contributed to the any match measure. Precision is poor with this combination for exact and boundary matches (P = 0.43) but improves for all matches (P = 0.66). Left boundary (P = 0.62), right boundary (P = 0.64), and non-boundary (R = 0.58) each contributed to the any match measure.

4.5.2.10 Evaluation 10

Corpus: PubMed abstracts Target: Treatment Dictionary: UMLS

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.42	0.44	0.43	0.43	0.40	0.41	0.43	0.38	0.40	0.43	0.38	0.40
Exact and left boundary matches	0.75	0.58	0.65	0.77	0.54	0.64	0.74	0.51	0.60	0.80	0.53	0.64
Exact and right boundary matches	0.76	0.58	0.66	0.79	0.54	0.64	0.77	0.52	0.62	0.81	0.53	0.64
Exact and non-boundary matches	0.58	0.52	0.55	0.59	0.47	0.52	0.58	0.45	0.51	0.61	0.46	0.53
All matches	0.79	0.59	0.68	0.82	0.55	0.66	0.80	0.53	0.64	0.83	0.54	0.65

Figure 4.16. Results from UMLS and PubMed treatment analysis.

UMLS dictionary and cTAKES pipeline combination shown in **Figure 4.16** is very poor for exact boundary matches (F = 0.43) but tremendously improves for all exact and any boundary match (F = 0.68). Recall is poor for exact matches (R = 0.42) but greatly improves for all matches (R = 0.79). Left boundary (R = 0.75), right boundary (R = 0.76), and non-boundary (R = 0.58) each contributed to the any match measure. Precision is poor with this combination for exact boundary matches (P = 0.44) but improves for any match (P = 0.59). The F-measure with UMLS dictionary and SKATE pipeline combination is very poor for exact boundary matches (F = 0.41) but improves for all exact and any boundary match (F = 0.66). Recall is poor for exact matches (R = 0.43) but tremendously improves for all matches (R = 0.82). Precision is poor with this combination for exact boundary matches (P = 0.40) but improves for any match (P = 0.55). Displayed in **Figure 4.16**, the F-measure with UMLS dictionary and NCBO Annotator pipeline combination is good for exact boundary matches (F = 0.40) but improves for all exact and any boundary match (F = 0.64). Recall is not good for exact matches (R = 0.43) but tremendously improves for all matches (R = 0.80). Precision is not good with this combination for exact and boundary matches (P = 0.38) but slightly improved for any match (P = 0.53). The F-measure with UMLS dictionary and MetaMap pipeline combination is very poor for exact boundary matches (F = 0.40) but improves for all exact and any boundary match (F = 0.65). Left boundary (F = 0.64), right boundary (F = 0.64), and non-boundary (F = 0.53) each contributed to the any match measure (F = 0.65). Recall is not good for exact matches (R = 0.43) but tremendously improves for all matches (R = 0.83). Left boundary (R = 0.80), right boundary (R = 0.81), and non-boundary (R = 0.61) each contributed to the any match measure. Precision is not good with

this combination for exact and boundary matches (P = 0.38) but slightly improves for any match (P = 0.54).

4.5.2.11 Evaluation 11

Corpus: Psychotherapeutic transcripts Target: Symptom Dictionary: PTSDO

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.74	0.83	0.79	0.80	0.84	0.82	0.74	0.80	0.77	0.76	0.82	0.78
Exact and left boundary matches	0.77	0.84	0.80	0.87	0.85	0.86	0.83	0.82	0.83	0.84	0.83	0.84
Exact and right boundary matches	0.76	0.84	0.80	0.83	0.84	0.84	0.77	0.81	0.79	0.78	0.82	0.80
Exact and non-boundary matches	0.75	0.83	0.79	0.80	0.84	0.82	0.74	0.80	0.77	0.76	0.82	0.79
All matches	0.78	0.84	0.81	0.89	0.85	0.87	0.86	0.83	0.84	0.87	0.84	0.85

Figure 4.17. Results from PTSDO and psychotherapeutic transcripts symptom analysis.

As seen in **Figure 4.17**, the F-measure with PTSDO dictionary and cTAKES pipeline combination is good for exact boundary matches (F = 0.79) and slightly improves for all exact and any boundary match (F = 0.81). Recall is respectable for exact matches (R = 0.74) but slightly improves for all matches (R = 0.78). Precision is good with this combination for exact boundary matches (P = 0.83). It only slightly improves for any match (P = 0.84). PTSDO-SKATE F-measure is great for exact boundary matches (F = 0.82) and is excellent for all exact and any boundary match (F = 0.87). Recall is good for exact matches (R = 0.80) but greatly improves for all matches (R = 0.89). Precision is great with this combination for exact boundary matches (P = 0.84). The F-measure with PTSDO dictionary and NCBO Annotator pipeline combination is good for exact boundary matches (F = 0.77) and improves for all exact and any boundary match (F = 0.84). Recall is good for exact matches (R = 0.74) but greatly improves for all matches (R = 0.86). The majority of the additional detection is recognized by the left boundary matches (R = 0.83). Precision is great with this combination for exact boundary matches (P = 0.80) improves for all matches (P = 0.83). PTSDO-MetaMap F-measure is good for exact boundary matches (F = 0.78) and is excellent for all exact and any boundary match (F = 0.85). Recall is good for exact matches (R = 0.76) but greatly improves for all matches (R = 0.87). Precision is great with this combination for exact boundary matches (P = 0.82).

4.5.2.12 Evaluation 12

Corpus: Psychotherapeutic transcripts Target: Symptom Dictionary: PILOTS

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.30	0.80	0.43	0.32	0.80	0.46	0.29	0.76	0.42	0.33	0.80	0.47
Exact and left boundary matches	0.32	0.81	0.46	0.36	0.82	0.50	0.31	0.77	0.44	0.37	0.82	0.51
Exact and right boundary matches	0.31	0.80	0.44	0.33	0.81	0.47	0.30	0.76	0.43	0.34	0.81	0.48
Exact and non-boundary matches	0.30	0.80	0.44	0.33	0.80	0.46	0.30	0.76	0.43	0.33	0.80	0.47
All matches	0.33	0.81	0.47	0.37	0.82	0.51	0.31	0.77	0.45	0.38	0.82	0.52

Figure 4.18. Results from PILOTS and psychotherapeutic transcripts symptom analysis.

Figure 4.18 displays that the F-measure with PILOTS dictionary and cTAKES pipeline combination is not good for exact boundary matches (F = 0.43) and only slightly improves for all exact and any boundary match (F = 0.47). Overall, recall is poor for exact matches (R = 0.30) and only slightly improves for all matches (R = 0.33). Precision is good with this combination for exact boundary matches (P = 0.80). The F-measure with PILOTS dictionary and SKATE pipeline combination is very poor for exact boundary matches (F = 0.46) and only slightly improves for all exact and any boundary match (F = 0.51). Recall is extremely poor for exact matches (R = 0.32) and only slightly improves for all matches (R = 0.37). Precision is good with this combination for exact boundary matches (P = 0.80) and improves for any match (P = 0.82). PILOTS-NCBO Annotator pipeline combination F-measure with is poor for exact boundary matches (F = 0.42) and only slightly improves for all exact and any boundary match (F = 0.45). Recall is extremely poor for exact matches (R = 0.29) and only slightly improves for all matches (R = 0.31). Precision is good with this combination for exact boundary matches (P = 0.76) and only slightly improves for any match (P = 0.77). The F-measure with PILOTS dictionary and MetaMap pipeline combination is poor for exact boundary matches (F = 0.47) and only slightly improves for all exact and any boundary match (F = 0.52). Recall is extremely poor for exact matches (R = 0.33) and only slightly improves for all matches (R = 0.38). Precision is good with this combination for exact boundary matches (P = 0.80) and only slightly improves for any match (P = 0.82).

4.5.2.13 Evaluation 13

Corpus: Psychotherapeutic transcripts Target: Symptom Dictionary: SNOMED-CT

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.52	0.78	0.63	0.55	0.79	0.65	0.49	0.73	0.59	0.58	0.77	0.66
Exact and left boundary matches	0.59	0.81	0.68	0.65	0.81	0.72	0.59	0.77	0.67	0.70	0.80	0.74
Exact and right boundary matches	0.56	0.79	0.65	0.59	0.80	0.68	0.54	0.75	0.63	0.63	0.78	0.70
Exact and non-boundary matches	0.56	0.80	0.66	0.59	0.80	0.68	0.53	0.75	0.62	0.62	0.78	0.69
All matches	0.62	0.81	0.71	0.66	0.82	0.73	0.66	0.79	0.72	0.76	0.81	0.78

Figure 4.19. Results from SNOMED-CT and psychotherapeutic transcripts symptom analysis.

The F-measure with SNOMED-CT dictionary and cTAKES pipeline combination displayed in **Figure 4.19** is not good for exact boundary matches (F = 0.63) but slightly improves for all exact and any boundary match (F = 0.71). Recall is poor for exact matches (R = 0.52) and only slightly improves for all matches (R = 0.62). Precision is good with this combination for exact boundary matches (P = 0.78). SNOMED-CT dictionary and SKATE pipeline F-measure is not good for exact boundary matches (F = 0.65) and slightly improves for all exact and any boundary match (F = 0.73). Recall is not good for exact matches (R = 0.55) but improves for all matches (R = 0.66). Precision is great with this combination for exact boundary matches (P = 0.79) and only slightly improves for any match (P = 0.82). SNOMED-CT dictionary and NCBO Annotator pipeline combination F-measure with is not good for exact boundary matches (F = 0.59) but greatly improves for all exact and any boundary match (F = 0.72). Recall is poor for exact matches (R = 0.49) but greatly improves for all matches (R = 0.66). Precision is good with this combination for exact boundary matches (P = 0.74) and improves for any match (P = 0.79). Shown in **Figure 4.19**, the F-measure with SNOMED-CT dictionary and MetaMap pipeline combination is not good for exact boundary matches (F = 0.66) but greatly improves for all exact and any boundary match (F = 0.78). Recall is poor for exact matches (R = 0.58) but greatly improves for all matches (R = 0.76). Precision is good with this combination for exact boundary matches (P = 0.77) and improves for any match (P = 0.81).

4.5.2.14 Evaluation 14

Corpus: Psychotherapeutic transcripts Target: Symptom Dictionary: NCI Thesaurus

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.51	0.78	0.62	0.54	0.78	0.64	0.49	0.73	0.58	0.59	0.79	0.67
Exact and left boundary matches	0.56	0.79	0.66	0.60	0.80	0.69	0.54	0.75	0.63	0.65	0.80	0.72
Exact and right boundary matches	0.54	0.79	0.64	0.57	0.79	0.66	0.52	0.74	0.61	0.63	0.80	0.70
Exact and non-boundary matches	0.52	0.78	0.62	0.54	0.78	0.64	0.49	0.73	0.59	0.59	0.79	0.68
All matches	0.57	0.80	0.66	0.61	0.80	0.69	0.57	0.76	0.65	0.69	0.81	0.74

Figure 4.20. Results from NCI Thesaurus and psychotherapeutic transcripts symptom analysis.

NCI Thesaurus and cTAKES pipeline combination F-measure is not good for exact boundary matches (F = 0.62) and only slightly improves for all exact and any boundary match (F = 0.66). Recall is poor for exact matches (R = 0.51) and only slightly improves for all matches (R = 0.57) as shown in **Figure 4.20**. Precision is good with this combination for exact boundary matches (P = 0.78). NCI Thesaurus and SKATE pipeline combination F-measure with is poor for exact boundary matches (F = 0.64). Recall is very poor for exact matches (R = 0.54) and slightly improves for all matches (R = 0.61). Precision is great with this combination for exact boundary matches (P = 0.78). The F-measure with NCI Thesaurus and NCBO Annotator is not good for exact boundary matches (F = 0.58) and slightly improves for all exact and any boundary match (F = 0.65). Recall is poor for exact matches (R = 0.49) but improves for all matches (R = 0.57). Precision is good with this combination for exact boundary matches (P = 0.73) and only slightly improves for all matches (P = 0.76). NCI Thesaurus-MetaMap pipeline combination F-measure is not good for exact boundary matches (F = 0.67) but is excellent for all exact and any boundary match (F = 0.74). Recall is poor for exact matches (R = 0.59) but improves for all matches (R = 0.69). Precision is good with this combination for exact boundary matches (P = 0.79).

4.5.2.15 Evaluation 15

Corpus: Psychotherapeutic transcripts Target: Symptom Dictionary: UMLS

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.56	0.76	0.65	0.59	0.77	0.67	0.59	0.76	0.67	0.62	0.76	0.68
Exact and left boundary matches	0.64	0.79	0.71	0.69	0.79	0.74	0.70	0.79	0.74	0.74	0.79	0.76
Exact and right boundary matches	0.59	0.77	0.67	0.64	0.78	0.70	0.64	0.77	0.70	0.66	0.77	0.71
Exact and non-boundary matches	0.59	0.77	0.67	0.63	0.78	0.70	0.63	0.77	0.69	0.65	0.77	0.71
All matches	0.65	0.79	0.71	0.71	0.80	0.75	0.71	0.79	0.75	0.80	0.80	0.80

Figure 4.21. Results from UMLS and psychotherapeutic transcripts symptom analysis.

UMLS and cTAKES pipeline F-measure is not good for exact boundary matches (F = 0.65) as seen in **Figure 4.21** and only slightly for all exact and any boundary match (F = 0.71). Recall is poor for exact matches (R = 0.56) and only slightly improves for all matches (R = 0.65). Precision is good with this combination for exact boundary matches (P = 0.76). The F-measure with UMLS dictionary and SKATE pipeline combination is not good for exact boundary matches (F = 0.67) but improves for all exact and any boundary match (F = 0.75). Recall is not good for exact matches (R = 0.59) but greatly improves for all matches (R = 0.71). Precision is good with this combination for exact boundary matches (P = 0.77). UMLS and NCBO Annotator F-measure with is poor for exact boundary matches (F = 0.67) but improves for all exact and any boundary match (F = 0.75). Recall is poor for exact matches (R = 0.59) but greatly improves for all matches (R = 0.71). Precision is good with this combination for exact boundary matches (P = 0.76) and only slightly improves for any match (P = 0.79). Shown in **Figure 4.21**, the F-measure with UMLS dictionary and MetaMap pipeline combination is not good for exact boundary matches (F = 0.68) but greatly improves for all exact and any boundary match (F = 0.80). Recall is poor for exact matches (R = 0.62) but greatly improves for all matches (R = 0.80). Precision is good with this combination for exact boundary matches (P = 0.76) and slightly improves for any match (P = 0.80).

4.5.2.16 Evaluation 16

Corpus: Psychotherapeutic transcripts Target: Treatment Dictionary: PTSDO

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.79	0.85	0.82	0.81	0.85	0.83	0.76	0.83	0.79	0.78	0.82	0.80
Exact and left boundary matches	0.86	0.86	0.86	0.91	0.86	0.89	0.84	0.84	0.84	0.84	0.83	0.83
Exact and right boundary matches	0.87	0.86	0.87	0.93	0.86	0.89	0.87	0.85	0.86	0.88	0.84	0.86
Exact and non-boundary matches	0.82	0.86	0.84	0.88	0.86	0.87	0.82	0.84	0.83	0.88	0.84	0.86
All matches	0.88	0.87	0.87	0.95	0.86	0.91	0.91	0.85	0.88	0.93	0.84	0.88

Figure 4.22. Results from PTSDO and psychotherapeutic transcripts treatment analysis.

The F-measure with PTSDO dictionary and cTAKES pipeline combination from **Figure 4.22** is very good for exact boundary matches (F = 0.82) and improves to very good for all exact and any boundary match (F = 0.87). Recall is good for exact matches (R = 0.79) and greatly improves for all matches (R = 0.88). Precision is great with this combination for exact boundary matches (P = 0.85). PTSDO-SKATE F-measure is very good for exact boundary matches (F = 0.83) and improves to tremendously good for all exact and any boundary match (F = 0.91). Overall, recall is good for exact matches (R = 0.81) and greatly improves for all matches (R = 0.95). Precision is great with this combination for exact boundary matches (P = 0.85). The F-measure with PTSDO dictionary and NCBO Annotator pipeline combination very good for exact boundary matches (F = 0.79) and greatly improves for all exact and any boundary match (F = 0.88). Recall is good for exact matches (R = 0.76) but greatly improves for all matches (R = 0.91). Precision is great with this combination for exact boundary matches (P = 0.83). PTSDO-MetaMap is very good for exact boundary matches (F = 0.80) and greatly improves for all exact and any boundary match (F = 0.88). Shown in **Figure 4.22**, recall is good for exact matches (R = 0.78) but greatly improves for all matches (R = 0.93). Precision is great with this combination for exact boundary matches (P = 0.82).

4.5.2.17 Evaluation 17

Corpus: Psychotherapeutic transcripts Target: Treatment Dictionary: PILOTS

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.32	0.67	0.43	0.33	0.64	0.43	0.35	0.64	0.45	0.35	0.62	0.45
Exact and left boundary matches	0.48	0.75	0.58	0.48	0.73	0.58	0.48	0.71	0.57	0.51	0.70	0.59
Exact and right boundary matches	0.39	0.71	0.50	0.40	0.69	0.51	0.39	0.66	0.49	0.42	0.66	0.51
Exact and non-boundary matches	0.39	0.71	0.50	0.41	0.69	0.51	0.38	0.66	0.49	0.43	0.67	0.52
All matches	0.58	0.78	0.66	0.59	0.77	0.67	0.57	0.74	0.64	0.61	0.74	0.67

Figure 4.23. Results from PILOTS and psychotherapeutic transcripts treatment analysis.

Displayed in **Figure 4.23**, the F-measure with PILOTS dictionary and cTAKES pipeline combination is not good for exact boundary matches (F = 0.43) but improves for all exact and any boundary match (F = 0.66). Recall is very poor for exact matches (R = 0.32) but improves for all matches (R = 0.58). Precision is not good with this combination for exact boundary matches (P = 0.67) but improves for any match (P = 0.78). PILOTS-SKATE F-measure is very poor for exact boundary matches (F = 0.43) but improves for all exact and any boundary match (F = 0.67). Recall is extremely poor for exact matches (R = 0.33) but improves for all matches (R = 0.59). Precision is decent with this combination for exact boundary matches (P = 0.64) and improves for any match (P = 0.77). PILOTS and NCBO Annotator F-measure is poor for exact boundary matches (F = 0.45) but improves for all exact and any boundary match (F = 0.64). Recall is very poor for exact matches (R = 0.35) but improves for all matches (R = 0.57). Precision is decent with this combination for exact boundary matches (P = 0.64) but improves for any match (P = 0.74). PILOTS-MetaMap F-measure shown in **Figure 4.23** is poor for exact boundary matches (F = 0.45) but improves for all exact and any boundary match (F = 0.67). Recall is very poor for exact matches (R = 0.35) but greatly improves for all matches (R = 0.61). Precision is not good with this combination for exact boundary matches (P = 0.62) but improves for any match (P = 0.74).

4.5.2.18 Evaluation 18

Corpus: Psychotherapeutic transcripts Target: Treatment Dictionary: SNOMED-CT

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.31	0.58	0.41	0.32	0.56	0.41	0.32	0.52	0.40	0.41	0.54	0.47
Exact and left boundary matches	0.43	0.65	0.52	0.43	0.63	0.51	0.43	0.59	0.50	0.60	0.64	0.62
Exact and right boundary matches	0.52	0.69	0.59	0.52	0.68	0.59	0.52	0.64	0.58	0.67	0.66	0.66
Exact and non-boundary matches	0.38	0.62	0.47	0.43	0.63	0.51	0.43	0.59	0.50	0.54	0.61	0.57
All matches	0.55	0.71	0.62	0.59	0.70	0.64	0.60	0.67	0.63	0.67	0.66	0.67

Figure 4.24. Results from SNOMED-CT and psychotherapeutic transcripts treatment analysis.

The F-measure with PILOTS dictionary and cTAKES pipeline combination shown in **Figure 4.24** is not good for exact boundary matches (F = 0.41) but improves for all exact and any boundary match (F = 0.62). Overall, recall is very poor for exact matches (R = 0.31) but improves for all matches (R = 0.55). Precision is poor with this combination for exact boundary matches (P = 0.58) but improves for any match (P = 0.71). The F-measure with PILOTS dictionary and SKATE pipeline combination is very poor for exact boundary matches (F = 0.41) but improves for all exact and any boundary match (F = 0.64). Recall is extremely poor for exact matches (R = 0.32) but improves for all matches (R = 0.59). Precision is not good with this combination for exact boundary matches (P = 0.56) but improves for any match (P = 0.70). PILOTS and NCBO Annotator pipeline combination is poor for exact boundary matches F-measure (F = 0.40) but improves for all exact and any boundary match (F = 0.63). Recall is very poor for exact matches (R = 0.32) but improves for all matches (R = 0.60). Displayed in **Figure 4.24**, precision is decent with this combination for exact boundary matches (P = 0.52) but improves for any match (P = 0.67). In the PILOTS dictionary and MetaMap pipeline, the F-measure is poor for exact boundary matches (F = 0.47) but improves for all exact and any boundary match (F = 0.67). As seen in **Figure 4.24**, recall is very poor for exact matches (R = 0.41) but greatly improves for all matches (R = 0.67). Precision is poor with this combination for exact boundary matches (P = 0.54) and only slightly improves for any match (P = 0.66).

4.5.2.19 Evaluation 19

Corpus: Psychotherapeutic transcripts Target: Treatment Dictionary: NCI Thesaurus

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.50	0.73	0.59	0.51	0.72	0.60	0.51	0.70	0.59	0.51	0.66	0.58
Exact and left boundary matches	0.79	0.81	0.80	0.80	0.80	0.80	0.80	0.78	0.79	0.89	0.77	0.83
Exact and right boundary matches	0.74	0.80	0.77	0.75	0.79	0.77	0.76	0.77	0.77	0.93	0.78	0.85
Exact and non-boundary matches	0.64	0.77	0.70	0.64	0.76	0.70	0.65	0.74	0.69	0.71	0.73	0.72
All matches	0.85	0.82	0.84	0.87	0.81	0.84	0.88	0.80	0.84	0.94	0.78	0.85

Figure 4.25. Results from NCI Thesaurus and psychotherapeutic transcripts treatment analysis.

PILOTS dictionary and cTAKES pipeline F-measure is decent for exact boundary matches (F = 0.59) but tremendously improves for all exact and any boundary match (F = 0.84) displayed in **Figure 4.25**. Recall is not good for exact matches (R = 0.50) but greatly improves for all matches (R = 0.85). Precision is good with this combination for exact boundary matches (P = 0.73) but improves for any match (P = 0.82). The F-measure with PILOTS dictionary and SKATE pipeline combination is decent for exact boundary matches (F = 0.60) but greatly improves for all exact and any boundary match (F = 0.84). Recall is good for exact matches (R = 0.51) but greatly improves for all matches (R = 0.87). Precision is good with this combination for exact boundary matches (P = 0.72) but improves for any match (P = 0.81). The F-measure with PILOTS dictionary and NCBO Annotator pipeline combination is decent for exact boundary matches (F = 0.59) but greatly improves for all exact and any boundary match (F = 0.84). Recall is not good for exact matches (R = 0.51) but tremendously improves for all matches (R = 0.88). Precision is good with this combination for exact boundary matches (P = 0.70) but improves for any match (P = 0.80). PILOTS-MetaMap F-measure from **Figure 4.25** is poor for exact boundary matches (F = 0.47) but improves for all exact and any boundary match (F = 0.67). Recall is not good for exact matches (R = 0.51) but tremendously improves for all matches (R = 0.94). Precision is not good with this combination for exact boundary matches (P = 0.66) but improves for any match (P = 0.78).

4.5.2.20 Evaluation 20

Corpus: Psychotherapeutic transcripts Target: Treatment Dictionary: UMLS

	cTAKES			SKATE			NCBO			MetaMap		
	R	P	F	R	P	F	R	P	F	R	P	F
Exact matches	0.67	0.70	0.69	0.69	0.70	0.70	0.67	0.66	0.67	0.71	0.64	0.67
Exact and left boundary matches	0.83	0.75	0.79	0.85	0.74	0.79	0.81	0.70	0.75	0.89	0.69	0.78
Exact and right boundary matches	0.87	0.76	0.81	0.93	0.76	0.83	0.90	0.72	0.80	0.95	0.70	0.81
Exact and non-boundary matches	0.76	0.73	0.74	0.82	0.73	0.77	0.80	0.70	0.74	0.84	0.68	0.75
All matches	0.90	0.76	0.83	0.93	0.76	0.83	0.90	0.72	0.80	0.95	0.70	0.81

Figure 4.26. Results from UMLS and psychotherapeutic transcripts treatment analysis.

The F-measure with UMLS dictionary and cTAKES pipeline combination is good for exact boundary matches (F = 0.69) but improves for all exact and any boundary match (F = 0.83). Displayed in **Figure 4.26**, recall is good for exact matches (R = 0.67) but greatly improves for all matches (R = 0.90). Precision is good with this combination for exact boundary matches (P = 0.70) and slightly improves for any match (P = 0.76). UMLS-SKATE combination F-measure is good for exact boundary matches (F = 0.70) but greatly improves for all exact and any boundary match (F = 0.83). Recall is good for exact matches (R = 0.69) but greatly improves for all matches (R = 0.93). Precision is good with this combination for exact boundary matches (P = 0.70) but improves for any match (P = 0.76). The F-measure with UMLS dictionary and NCBO Annotator pipeline combination is decent for exact boundary matches (F = 0.67) but greatly improves for all exact and any boundary match (F = 0.80). Recall is decent for exact matches (R = 0.67) but tremendously improves for all matches (R = 0.90). Precision is good with this combination for exact boundary matches (P = 0.66) but improves for any match (P = 0.72). The F-measure with UMLS dictionary and MetaMap pipeline combination is decent for exact boundary matches (F = 0.67) but greatly improves for all exact and any boundary match (F = 0.81). Recall is good for exact matches (R = 0.71) but tremendously improves for all matches (R = 0.95). Precision is not good with this combination for exact boundary matches (P = 0.64) but improves for any match (P = 0.70).

4.5.3 Statistical Analysis

A Kruskal-Wallis test is conducted to determine whether F-measures obtained on identification of symptoms and treatments from PubMed case reports are different for twenty

combinations of four text mining pipelines with five reference terminological resources with n equal to the number of documents analyzed: (a) cTAKES-PTSDO (n=122); (b) SKATE-PTSDO (n=122); and (c) NCBO-PTSDO (n=122); (d) MetaMap-PTSDO (n=122); (e) cTAKES-PILOTS (n=122); (f) SKATE-PILOTS (n=122); and (g) NCBO-PILOTS (n=122); (h) MetaMap-PILOTS (n=122); (i) cTAKES-SNOMEDCT (n=122); (j) SKATE- SNOMEDCT (n=122); and (k) NCBO-SNOMEDCT (n=122); (l) MetaMap-SNOMEDCT (n = 90); (m) cTAKES-NCIT (n=122); (n) SKATE-NCIT (n=122); and (o) NCBO-NCIT (n=122); (p) MetaMap-NCIT (n=122); (q) cTAKES-UMLS (n=122); (u) SKATE-UMLS (n=122); and (r) NCBO-UMLS (n=122); (s) MetaMap-UMLS (n=122). For post-hoc analysis, a Dunn's pairwise comparison with Bonferroni correction determined F-measure statistical differences among each text mining pipeline and terminology resource combination.

For PubMed case reports, a Kruskal-Wallis test shows that there is a statistically significant difference in F-measure symptom identification between the twenty groups, $\chi^2(19) = 370.342$, $p = 0.0001$. Post-hoc comparison finds that a total of 60 significant differences occur among the 190 pipeline and terminology groupings. **Table 4.7** highlights the significant differences for PTSDO and the other terminologies analyzed with each pipeline for symptoms in PubMed case reports. Interestingly, F-measures obtained with PTSDO for cTAKES and SKATE pipelines are significantly different ($p<0.00026$) among each dictionary other than PILOTS. When supporting the NCBO Annotator, PTSDO is significantly different ($p<0.00026$) than the NCI-Thesaurus and the UMLS and has a mean F-measure difference of 0.12. Relative to PTSDO, the average effect is determined to be 0.1518 meaning 15.2% of the variability in the F-measure score is accounted for by the dictionary that supports the pipeline which is a decent impact in the mental health text processing domain. Analyzing PTSDO with MetaMap, it is only significantly different ($p<0.00026$) from the NCI-Thesaurus with a 0.11 mean F-measure difference. For symptom identification in PubMed case reports, PTSDO supported the highest accuracy across each of the text mining platforms.

Table 4.7. Significant difference (p<0.00026) between PTSDO and other terminologies

Corpus: PubMed case reports, **Target:** symptom

	PTSDO			
	cTAKES	SKATE	NCBO	MetaMap
PILOTS				
SNOMED-CT	✓	✓		
NCIT	✓	✓	✓	✓
UMLS	✓	✓	✓	

A Kruskal-Wallis test shows that there is a statistically significant difference in F-measure treatment identification between the twenty groups on PubMed case reports, $\chi^2(19) = 124.014$, $p = 0.0001$. Post-hoc comparison finds a total of 14 significant differences occurring among the 190 pipeline and terminology groupings. **Table 4.8** highlights the significant differences for PTSDO and the other terminologies analyzed with each pipeline for treatments in PubMed case reports. The only pipeline supported with PTSDO that produces a significant difference is SKATE. PTSDO is significantly different ($p<0.00026$) from PILOTS, SNOMED-CT, and the UMLS. Relative to PTSDO, the average effect is determined to be 0.051 meaning 5.1% of the variability in the F-measure score is accounted for by the dictionary that supports the pipeline. For treatment identification in PubMed case reports, PTSDO supported the highest accuracy across each of the text mining platforms. The SKATE text mining platform, overall, exhibits a better performance than the other pipelines.

Table 4.8. Significant difference (p<0.00026) between PTSDO and other terminologies

Corpus: PubMed case reports, **Target:** treatment

	PTSDO			
	cTAKES	SKATE	NCBO	MetaMap
PILOTS		✓		
SNOMED-CT		✓		
NCIT				
UMLS		✓		

A Kruskal-Wallis test is conducted to determine whether F-measures obtained on identification of symptoms and treatments from psychotherapeutic transcripts is significantly

different for twenty combinations of four text mining pipelines with five reference terminological resources with n equal to the number of documents analyzed: (a) cTAKES-PTSDO (n=82); (b) SKATE-PTSDO (n=82); and (c) NCBO-PTSDO (n=82); (d) MetaMap-PTSDO (n=82); (e) cTAKES-PILOTS (n=82); (f) SKATE-PILOTS (n=82); and (g) NCBO-PILOTS (n=82); (h) MetaMap-PILOTS (n=82); (i) cTAKES-SNOMEDCT (n=82); (j) SKATE-SNOMEDCT (n=82); and (k) NCBO-SNOMEDCT (n=82); (l) MetaMap-SNOMEDCT (n=82); (m) cTAKES-NCIT (n=82); (n) SKATE-NCIT (n=82); and (o) NCBO-NCIT n=82); (p) MetaMap-NCIT (n=82); (q) cTAKES-UMLS (n=82); (u) SKATE-UMLS (n=82); and (r) NCBO-UMLS (n=82); (s) MetaMap-UMLS (n=82). For post-hoc analysis, a Dunn's pairwise comparison with Bonferroni correction determines F-measure statistical differences among each text mining pipeline and terminology resource combination.

A Kruskal-Wallis test shows that there is a statistically significant difference in F-measure symptom identification between the twenty groups on psychotherapeutic transcripts, $\chi^2(19) = 809.265$, $p = 0.0001$. Post-hoc comparison finds a total of 96 significant differences occurring among the 190 pipeline and terminology groupings. **Table 4.9** highlights the significant differences for PTSDO and the other terminologies analyzed with each pipeline for symptoms in psychotherapeutic transcripts. Interestingly, F-measures obtained with PTSDO for cTAKES, SKATE, and NCBO Annotator pipelines are significantly different ($p<0.00026$) among each of the dictionaries and has a mean F-measure differences of 0.07. For symptom identification in psychotherapeutic transcripts, PTSDO supported the highest accuracy across each of the text mining platforms. Relative to PTSDO, the average effect is determined to be 0.2938 meaning 29.4% of the variability in the F-measure score is accounted for by the dictionary that supports the pipeline. With MetaMap, PTSDO is significantly different ($p<0.00026$) from both PILOTS and NCI Thesaurus with a 0.15 mean F-measure difference.

4.5.4 Dictionaries and Pipelines

PTSDO produces the highest F-measure over each pipeline implementation on symptom concepts from PubMed case reports. With the pipeline SKATE, it produced the highest accuracy metrics (F=0.94; R=0.92; P=0.95). On this corpus, the worst performing terminology is the NCI-Thesaurus with supporting the cTAKES pipeline (F=0.69; R=0.63; P=0.77). Additionally, treatment concepts from the case reports extracted with PTSDO obtained the highest F-measure

(F=0.82; R=0.83; P=0.81) with the SKATE pipeline, while PILOTS produced the lowest (F=0.52; R=0.42; P=0.68) with the cTAKES pipeline. The coverage of the terminology resources varies in granularity as displayed in **Tables 4.27a** and **4.27b**. SKATE text mining pipeline implementation is displayed due to it supporting the highest accuracy over each of the dictionaries. For PubMed symptom matches with SKATE, there is little difference in overall F-measure but a greater difference for treatment identification. The majority of the dictionary and pipeline implementations is improved by the exact and left boundary matches. Locating any match will likely be useful for many use cases as the entity recognition can point the clinician or researcher back to the identified sentence for further investigation.

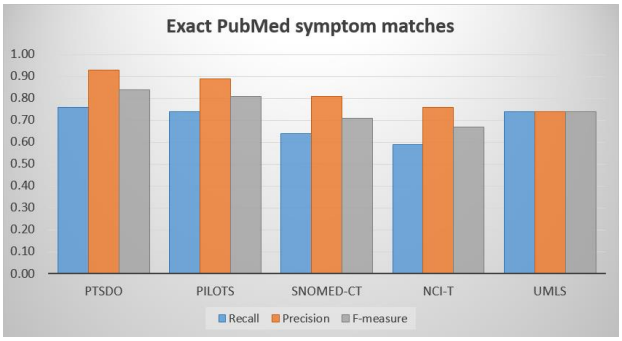


Figure 4.27a. Symptom metrics of dictionaries supporting SKATE

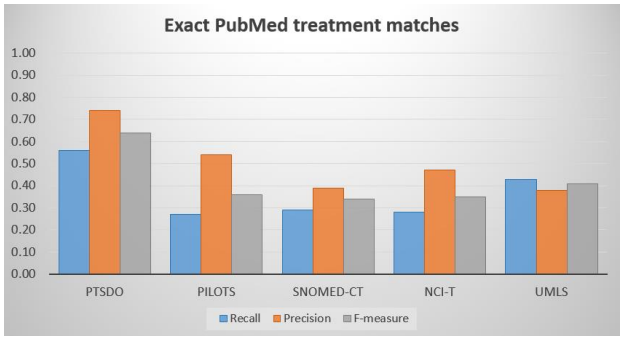


Figure 4.27b. Treatment metrics of dictionaries supporting SKATE

PTSDO produces the highest F-measure (F=0.87; R=0.89; P=0.85) with the SKATE pipeline on symptom concepts from psychotherapeutic transcripts. The worst performing terminology on this corpus is PILOTS (F=0.47; R=0.33; P=0.81) when supporting cTAKES. Treatment concepts from the case reports extracted with PTSDO supporting the SKATE pipeline obtains the highest F-measure (F=0.91; R=0.95; P=0.86), while SNOMED-CT produces the lowest (F=0.62; R=0.55; P=0.71) when supporting cTAKES. **Tables 4.28a** and **4.28b** display an average F-measure with terminologies support of psychotherapeutic transcript entity identification with SKATE. Precision displays little variability between terminology resources while recall is much improved by support of PTSDO.

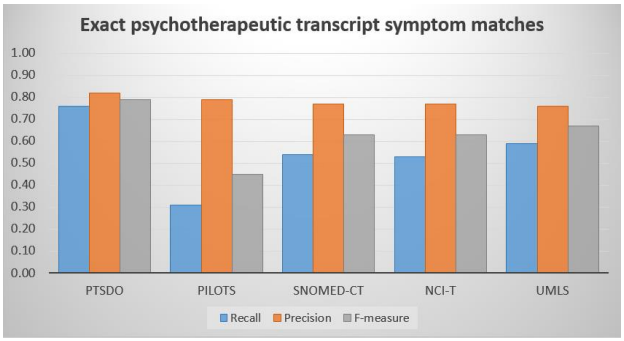


Figure 4.28a. Symptom metrics of dictionaries supporting SKATE

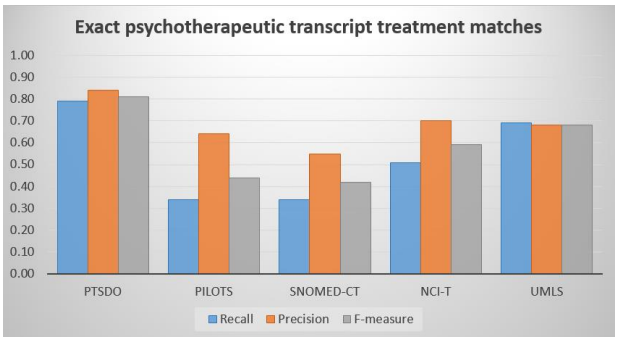


Figure 4.28b. Treatment metrics of dictionaries supporting SKATE

Table 4.9. Significant difference ($p<0.00026$) between PTSDO and other terminologies

Corpus: Psychotherapeutic transcripts, **Target:** symptom

	PTSDO			
	cTAKES	SKATE	NCBO	MetaMap
PILOTS	✓	✓	✓	✓
SNOMED-CT	✓	✓	✓	
NCIT	✓	✓	✓	✓
UMLS	✓	✓	✓	

A Kruskal-Wallis test shows that there is a statistically significant difference in F-measure symptom identification between the twenty groups on psychotherapeutic transcripts, $\chi^2(19) = 518.823$, $p = 0.0001$. Post-hoc comparison finds a total of 79 significant differences occur among the 190 pipeline and terminology groupings. **Table 4.10** highlights the significant differences for PTSDO and the other terminologies analyzed with each pipeline for treatments in psychotherapeutic transcripts. For treatment identification in psychotherapeutic transcripts, PTSDO supported the highest accuracy across each of the text mining platforms. F-measures obtained with PTSDO for NCBO Annotator pipeline are significantly different ($p<0.00026$) among each of the dictionaries. When supporting cTAKES, SKATE, and MetaMap, PTSDO is significantly different ($p<0.00026$) from PILOTS and SNOMED-CT with each pipeline. Relative to PTSDO, the average effect is determined to be 0.3165 meaning 31.7% of the variability in the F-measure score is accounted for by the dictionary that supports the pipeline.

Table 4.10. Significant difference (p<0.00026) between PTSDO and other terminologies

Corpus: Psychotherapeutic transcripts, **Target:** treatment

	PTSDO			
	cTAKES	SKATE	NCBO	MetaMap
PILOTS	✓	✓	✓	✓
SNOMED-CT	✓	✓	✓	✓
NCIT			✓	
UMLS			✓	

The results show several statistical differences with PTSDO and the other terminology resources. The increase in metrics from exact matching to partial matching highlights the importance of boundary identification.

4.6 Discussion

Understanding the narrative text available in the PTSD domain is critical to the success of future clinical decision support applications. Extracting the entities within this disorder by clinicians, curators, researchers, and other stakeholders takes considerable time and resources. Text mining initiatives ease these processes and improves performance of developing these clinical applications. Development of gold standards are necessary for the training and evaluation of text mining and NLP pipelines. This evaluation and subsequent error analysis identifies problems associated with the terminology, the pipeline, or in the creation of the gold standard itself. Identifying the errors is an important step to improving system accuracy and to understanding the strengths and limitations of the overall approach to entity identification.

4.6.1 Gold Standard Development

There is much ambiguous text in natural language resulting in time consuming and laborious tasks in the annotation process. The personnel availability of annotators and domain experts is a major limitation in this task. However, it is a feasible undertaking despite the difficulties the annotators did experience. The inter-annotator agreements obtained from the symptom and treatment gold standard creation in both corpora are reasonably high. The PubMed abstracts IAA average F-measure of 0.90 and the psychotherapeutic transcript average F-measure of 0.83 in

comparison is respectable to IAAs calculated in prior i2b2 annotation tasks. For example, the ground truth generated by the community obtained F-measures above 0.90 against the ground truth of the experts for the i2b2 medication challenge [344]. The IAA from this research is also comparable to the average F-measure of 0.82 for participating systems in the 2014 i2b2 annotation task of risk factors for heart disease in clinical narratives for diabetic patients [345]. It is very possible that annotators without the background experience of those in this gold standard development would not achieved as high an agreement. Without experts performing tasks, many of the limitations could be improved by more comprehensive annotation guidelines as well as more time for annotation training. There is also an opportunity to develop more robust annotation tools that are much more user friendly, specifically for identification of named entities.

Some of the most difficult tasks in this annotation process is building the schema and guidelines. Before the start of this project, stakeholders were asked to agree upon the definitions and granularity needs of the systems the annotated corpora will be supporting. The requirements depend upon the clinical questions that need to be answered in order to determine the appropriate level of information to be annotated. This gold standard creation was hampered when changing stakeholder requirements for the terminology development affected the annotation schema and tasks. Reasonable modifications to requirements was allowed under agile methodology, however once the gold standard annotation guide was finalized, adjustments could not be allowed.

Annotation of signs/symptoms related concepts is perceived as the most difficult and task due to ambiguous lexical variants and the vast alternative portrayals of synonymous text. Accurate annotation of symptom and treatment concepts in the psychotherapeutic transcripts, and to some degree in the PubMed case reports, is inhibited by existing concepts that pertain to the mental health domain, but not explicitly to the sub-domain of PTSD. In addition to the guidelines, annotators were able to reference the UMLS Terminology Services (UTS) and its domain knowledge which was helpful but did slow down the process as a whole. However, accuracy and the achievement of the overall high inter-annotator agreement is paramount to the subsequent tasks and experiments in the dissertation. There is a definite cost and benefit analysis to evaluate when annotators have a tool to reference versus relying solely on the annotation guidelines and their respective user knowledge.

A majority of the referential concepts did not match PTSD-related symptom and treatment concepts in the annotated corpus. This is not an unexpected finding as the annotators were briefed that the goal of this research is to enhance existing resources such as the UMLS for the coverage of PTSD. Many concepts are medically generic and do not contain enough detail to include synonymous terms in order to guide annotation for the disorder's domain entities. Providing a large number of examples on hand is identified by the annotators as the most beneficial item. Many of the notes and examples the annotators created themselves contributed to the improvement of the guidelines and can be constructive to future annotation projects.

The preliminary inter-annotator agreements improved for each annotator after every practice session. Overall, agreement of both symptom and treatment concepts is higher in the PubMed abstracts compared to the psychotherapeutic transcripts. Inter-annotator agreement is, on the average, higher for treatment concepts in both corpora. More fully developed guidelines could improve both entities and corpora IAA. The psychotherapeutic transcripts' vague language, lengthiness, and "rambling patient thoughts" were reported to affect the annotations as well. However, excellent agreement for all annotation tasks was able to be reached attributable to the annotators expertise and input into guidelines. While this corpora of annotated data is difficult to build, a benefit is that it is easily shareable to a multitude of text mining tasks and its methods can support other biomedical applications.

4.6.2 Error Analysis

The types and reasons for false-negative errors vary greatly and consist of the following categories: a) not available in dictionary; b) pipeline deficiencies; c) annotation mistakes; d) textual misidentification; and e) semantic type errors. The majority of false-negatives is due to concepts simply not existing in a dictionary, thus are not identified due to no available reference for the dictionary-lookup algorithm. Another error category is text mining deficiencies such as tokenization, stemming, normalization, and word sense disambiguation, etc. that causes inaccuracies beyond the control of the terminology reference. The next category is annotation corpus mistakes which is due to entities that should not have been annotated in the gold standard. Errors also derive from textual misidentification such as misspellings in the text, negation errors, and missed identification due to acronyms and abbreviations. Lastly, false-negatives are identified from the semantic type restriction that is placed on available concepts from the

terminologies implemented. These semantic types are removed in order to reduce the high number of false-positive errors, however, this limitation does produce false-negatives.

Table 4.11 displays a summary of false-negative errors from exact and all partial matches of symptoms from PubMed case reports. The average false-negatives across each dictionary resource is presented for each error type category. The majority of errors for identifying PubMed symptoms is due to no available reference in the dictionaries referenced. This error can be sub-divided into "dictionary missing concept" where it does not exist in a terminology resource and "dictionary missing synonym phase" where explicitly matching string representation is missing from the source. An example of the latter subtype in the table shows, for several dictionary and pipeline combinations, 'fear of open places' does not have a mapping to the concept of 'agoraphobia.' There are also substantial misidentifications due to textual recognition errors and many due to pipeline shortcomings. There are very few annotations found in the gold standard that should not have been annotated. Semantic types are restricted which does produce identification errors. An example shown in the table is the concept of 'war' that is not identified due to the non-inclusion of its semantic type Activity.

Table 4.11. False-negative symptom error analysis from PubMed case reports

	Example		Number of
Type of error	**Text**	**Concept missed**	**cases (%)**
a) Not available in dictionary	*...relaxation techniques to treat fear of open places persisting...*	agoraphobia	192 (37.5%)
b) Pipeline deficiencies	*...-related pictures correlated with re-experiencing and dissociation.*	re-experience	109 (21.2%)
c) Annotation mistakes	*... intolerance uncertainty and low tolerance of emotional distress.*	uncertainty	11 (2.1%)
d) Textual misidentification	*...between hyper-arousal symptoms and response within the gray matter.*	hyperarousal	127 (24.8%)
e) Semantic type error	*...effects of exposure to war, conflict and terrorism on young children.*	war	74 (14.4%)

False-positive categories consist of the following: a) incorrect sub-domain; b) pipeline deficiencies; and c) annotation mistakes. Many semantic types are restricted to reduce errors, but there are still many identified entities outside the domain of PTSD. While the non-inclusion of specific semantic types assists in increasing precision by ignoring irrelevant concepts within the domain, removing too many can have unintended consequences. It is not possible to restrict semantic types to alleviate all potential errors and over restriction would lead to misidentification of relevant PTSD concepts. Next, some of the same text mining pipeline deficiencies that cause false-negatives are also responsible for false-positives. Lastly, gold standard creation errors

missed annotation of some entities that should have been labeled as a relevant symptom or treatment concept for the domain.

Table 4.12 displays a summary of false-positive errors from exact and all partial matches of PubMed case report symptoms. The average false-positive errors across each dictionary resource is presented for each error type category. The majority of false-positives on this corpus is due to text-mining pipeline errors. Many of the pipelines recognized only partial concepts due to tokenization deficiencies such as in only identifying 'fight' in the concept of 'fight-or-flight.' Very few errors are due to annotation mistakes. In the example below, 'anhedonia' is correctly identified, however counted as a false-positive due to its lack of inclusion in the gold standard.

Table 4.12. False-positive symptom error analysis from PubMed case reports

	Example		Number of
Type of error	**Text**	**Concept identified**	**cases (%)**
a) Incorrect sub-domain	*…antipsychotic medication for treatment-resistant major depressive disorder.*	depressive	113 (29.5%)
b) Pipeline deficiencies	*The classic fight-or-flight response to perceived threat…*	fight	263 (68.7%)
c) Annotation mistakes	*The findings provide support for the DSM-5, anhedonia and hybrid models…*	anhedonia	7 (1.8%)

A summary of false-negative errors from exact and all partial matches of treatments from PubMed case reports is shown in **Table 4.13**. The average false-negatives across each dictionary resource is presented for each error type category. A slight majority of errors on this corpus is due to text mining deficiencies (29.8%) followed by textual identification errors (28.6%). Missing concepts within dictionaries averages to a count of 74 between terminology resources. Intellectual Product is a semantic type not included for symptom identification, however the entity 'image' is assigned this type, thus missed and counted as an error.

Table 4.13. False-negative treatment error analysis from PubMed case reports

	Example		Number of
Type of error	**Text**	**Concept missed**	**cases (%)**
a) Not available in dictionary	*…enhanced magnetic resonance imaging to measure hippocampi…*	magnetic resonance imaging	74 (28.3%)
b) Pipeline deficiencies	*…provided comprehensive psycho educational reporting to parents.*	psychoeducation	78 (29.8%)
c) Annotation mistakes	*…finding ways to self-medicate the headaches for the patient.*	self-medicate	4 (1.6%)
d) Textual misidentification	*To determine the safety and efficacy of EMDR in adult refugees…*	EMDR	75 (28.6%)
e) Semantic type error	*…to address the repeated images of the experience…*	image	31 (11.7%)

In **Table 4.14**, a summary displays the exact and all partial match false-positive errors of treatments from PubMed case reports. The average false-positive percentages across each dictionary resource is presented for each error type category. Again, a great majority of errors is due to pipeline deficiencies with a large portion due to identification of concepts not applicable to the PTSD domain. A caveat to the example given, 'methylphenidate' has recently been explored in the treatment of the disorder, however there are no conclusive findings to include the medication in the gold standard. There are minimal annotation mistakes such as the many treatment acronyms unidentifiable by annotators.

Table 4.14. False-positive treatment error analysis from PubMed case reports

	Example		Number of
Type of error	**Text**	**Concept identified**	**cases (%)**
a) Incorrect sub-domain	*...cases of apathy treated with a regimen of methylphenidate...*	methylphenidate	130 (41.3%)
b) Pipeline deficiencies	*...altering destructive self-talk, taking time-out to practice...*	time-out approach	180 (57.3%)
c) Annotation mistakes	*Key results for successful CISM were accessibility of CIPS team...*	CISM	4 (1.4%)

Table 4.15 displays a summary of false-negative errors from psychotherapeutic transcript symptoms. The average false-negatives across each dictionary resource is presented for each error type category. Almost half of these errors in this corpus are due to missing dictionary references. In this example, the synonym 'hyperalertness' for the concept 'hypervigilance' is not available in the terminology resource. 'Delusion' should not have been annotated which was a mistake made during the gold standard creation. A textual misidentification is highlighted by the fact that 'diminished interest' did not map to 'loss of interest.' For the semantic type error, 'difficulty' is recognized but 'physical contact' is of type Phenomenon or Process which is not included.

Table 4.15. False-negative symptom error analysis from psychotherapeutic transcripts

	Example		Number of
Type of error	**Text**	**Concept missed**	**cases (%)**
a) Not available in dictionary	*...sensed the hyperalertness of my mind and the reactions from..*	hypervigilance	1403 (46.4%)
b) Pipeline deficiencies	*...perceived from the impact-traumatized from the occurrence.*	traumatic	553 (18.3%)
c) Annotation mistakes	*Family members under the delusion of grandeur found among...*	delusion	169 (5.6%)
d) Textual misidentification	*He experienced a diminished interest in activity.*	loss of interest	529 (17.5%)
e) Semantic type error	*...comes from experiencing physical contact difficulty at times.*	physical contact difficulty	369 (12.2%)

A summary of exact and all partial match symptom false-positive errors from psychotherapeutic transcripts is shown in **Table 4.16**. The average false-positive percentages across dictionary resources is presented for each error type category. 57.4% of errors are due to pipeline deficiencies and 34.8% are due to inaccurate domain concept recognition. The identified concept of 'seizure' is not a concept explicit to the domain of PTSD. The text mining pipeline interpreted the numerical reference to 'number' and mapped it to the concept of 'numb.' 'Racing thought' is a domain symptom that is missed in the gold standard development.

Table 4.16. False-positive symptom error analysis from psychotherapeutic transcripts

	Example		Number of
Type of error	**Text**	**Concept identified**	**cases (%)**
a) Incorrect sub-domain	*... including cerebral malaria, sleep deprivation seizures, and toxic…*	seizure	437 (34.8%)
b) Pipeline deficiencies	*...a growing number of days missed from school is starting to…*	numb	721 (57.4%)
c) Annotation mistakes	*...such as flight of ideas, racing thought, and grandiose ideation.*	racing thought	98 (7.8%)

In **Table 4.17**, a false-negative error summary from exact and all partial matches of psychotherapeutic transcript treatments is displayed. False-negatives error averages across each dictionary resource is presented for each error type category. 36.2% of errors are due to missing concepts in a dictionary such as the treatment task of asking a 'challenging question.' Pipeline deficiencies account for 27.1% of errors and textual misidentification made up 22.3% as shown in the missed acronym of 'transcendental meditation' example. Annotation mistakes and missing semantic type errors are minimal.

Table 4.17. False-negative treatment error analysis from psychotherapeutic transcripts

	Example		Number of
Type of error	**Text**	**Concept missed**	**cases (%)**
a) Not available in dictionary	*...challenging questions were reviewed in order to understand…*	challenging question	129 (36.2%)
b) Pipeline deficiencies	*... behaviors and encouraging the development of self-caring routines.*	self-care behavior	97 (27.1%)
c) Annotation mistakes	*...information on general patterns of self-regulation failure.*	self-regulation	18 (5.1%)
d) Textual misidentification	*...practicing TM has been ritualistic in my routine.*	transcendental meditation	79 (22.3%)
e) Semantic type error	*...development of positive relational patterns that precedes…*	relational pattern	33 (9.3%)

Displayed in **Table 4.18**, a summary of psychotherapeutic transcript false-positive errors from exact and all partial matches of treatments is described. The average false-positives across

each dictionary resource is presented for each error type category. The term 'information' is too vague to be correct in the first example below. The percentage of incorrect domain and pipeline deficiency errors are relatively equal at around 40%. Shown in the second example is a negation miss, which is one of the most common pipeline deficiencies. The treatment concept of 'mindfulness,' which is a therapeutic mental process, should have been annotated in the gold standard.

Table 4.18. False-positive treatment error analysis from psychotherapeutic transcripts

	Example		Number of
Type of error	**Text**	**Concept identified**	**cases (%)**
a) Incorrect sub-domain	*...revisited information from process for employment...*	informational process	179 (46.6%)
b) Pipeline deficiencies	*...not encouragingly provided support to needs...*	provide encouragement	168 (43.6%)
c) Annotation mistakes	*...to achieve complete mindfulness of current circumstances.*	mindfulness	38 (9.8%)

Errors produced from lack of dictionary references are correctable with the methods described within this dissertation. Pipeline deficiencies are also correctable however, every system will typically be prone to some NLP shortcomings depending upon the needed tasks. Spelling mistakes are found such as 'anzer' unrecognizable for the concept 'anger.' These errors are unavoidable and inherent in most narrative or transcribed text. Acronyms and abbreviations are prevalent in the domain such as 'SI' for the 'suicidal ideation' concept. These will continue to be a constraint as the entire mental health domain expands. Other sources of textual errors include incorrect assignment of part-of-speech tags and/or incorrect chunking and parsing which cause the look-up algorithm failures. The task of keeping up with comprehensive knowledge coverage resources will continue to be a time-consuming limitation for maintaining accuracy. As new knowledge is inserted within the domain, it must be identified and included in the terminology. The more automation that can be implemented into these tasks, the more researchers and developers can maintain efficiency.

Dictionary-based terminology resources do not perform equally on corpora with text processing pipelines. Many existing terminologies are missing PTSD-related concepts and terms which do not support dictionary-lookup features for identification. Additionally, the non-specificity in the terminologies fosters the identification of spurious concepts not applicable to the domain of the disorder. The accuracy of dictionary coverage varies significantly between many of the text processing pipelines. This approach is particularly useful as the first step in

practical information extraction from clinical resources. PTSDO is example of improved coverage for a specific sub-domain in mental health, however there are continued improvements to be made. Its implementation displays excellent accuracy in support of the best of breed text mining pipelines and in comparison, to some of the most robust terminology sources. Expanded application can make inroads into information extraction, question answering, parsing, machine translation, and providing the framework for metadata to support the Semantic Web.

4.7 Conclusion

This research describes the development of a symptom and treatment concept gold standards for PubMed abstracts and psychotherapeutic transcripts, both with excellent inter-annotator agreement. The evaluation of four text mining pipelines in combination with five terminology resources was performed on these newly developed gold standards to include statistical analysis for significant difference. The results display the complexity of performing PTSD symptom and treatment entity identification. Shown is the importance of a domain-specific terminology resource with sufficient coverage for greater recall, and limiting semantic type capability for improved precision. Greater attention is needed for identification of appropriate concepts with explicit meaning through word sense disambiguation processes. By limiting semantic types in this research, it is an effective quick fix to improve disambiguation accuracy. Encompassing complete word sense disambiguation is beyond the scope of this dissertation, but complementary studies regarding similar tasks to this research would recognize benefits from its further considerations.

It is found that respectable accuracy metrics can be obtained both with utilizing existing resources and developing new systems. It is recommended to make use of these vetted terminology resources as a base to support the development of future vocabulary initiatives. Text mining pipelines, such as cTAKES, that support the processing of natural language text but also incorporate the input of progressing technologies such as clinical document architecture (CDA) will advance similar research initiatives tremendously. The largest deficiency is terminology availability limitations and the results show that attention to this area of clinical research can greatly improve future text mining and biomedical applications. The continued synthesis can expand upon the identification of relevant information by mining text and transforming it into structured data with salient categorization and meta-data properties. These detailed extraction

capabilities will be necessary to future implementations of clinical decision support. While continuing to improve functional competencies, considerations to system usability will be vital for wide-spread acceptance.

Chapter 5

Clustering and affinity analysis of PTSD signs and symptomatology derived from a hybrid taxonomic and pattern recognition approach to concept mining

5.1 Introduction

Commonly discussed is the need for initiatives and tools to support big data management strategies and effective solutions to combat information overload, however, it is often understated that we are still in the early stages data accumulation making future technology ingenuities as data continues to grow infinitely [245, 307-310]. Healthcare data can be collected from various sources such as literature, hospitals, etc. The data can be used to analyze patterns present in the data to further support retrieving information about various diseases and disorders giving medical researchers the ability to update and analyze changes in patterns of data [220]. A portion of biomedical and clinical narrative require professionals to make educated guesses on textual meaning. Applying rigorous algorithms and data analytical techniques support more informed decisions. These advanced practices support future precision medicine approaches as well as lay the framework for clinical decision support. Future research can expand to create more user-friendly systems with modular development to allow component specialization. The results are likely not to be one hundred percent correct but the process allows researcher, analysts, and scientists to refocus the analyses of the data with alternative algorithms, tools, techniques, and approaches that portray the best meaning [145, 189, 219].

5.1.1 Data Mining

Data mining is a technique for extracting and discovering implicit, previously unknown, and potentially useful information from collections of data [344]. A large database represents a great amount of information which can be potentially very useful if extracted and summarized correctly and from the right point of view. Using statistical tools and modeling techniques, one can discover interesting and hidden patterns in the data. These patterns may not be easily detected using traditional methods. Data mining technology is currently being used by a number

of industries, including pharmaceutical and biomedicine [345-346]. The decreasing cost of electronic data storage has made it economically feasible to collect and maintain large, long-term databases. The data mining technique draws ideas from a number of disciplines such as statistics, machine learning and database administration systems. Ideas from these disciplines are used to characterize the information which may be extracted from a large database and to quantify the usefulness of this information [347-348].

5.1.2 Clustering of Healthcare Data

According to a highly cited background by Jain et al., clustering is the "unsupervised classification of data points or feature vectors into groups of constellations" based upon user selected characteristics of the data [207]. First used by Tryon [206] in 1939, the term cluster analysis encompasses a number of different algorithms and methods for grouping objects of similar kind into respective categories. Unlike many other statistical procedures, cluster analysis methods are mostly used when an a priori hypotheses is not available, and researchers are in the exploratory phase. There are over 100 published clustering algorithms and this data mining technique has been useful to healthcare researchers in many contexts. In 1975, Hartigan discussed a thorough set of summaries of the many published studies reporting the results of cluster analyses. For example, in the field of medicine, clustering diseases, cures for diseases, or symptoms of diseases can lead to very useful taxonomies. In the field of psychiatry, the correct diagnosis of clusters of symptoms such as paranoia, schizophrenia, etc. is essential for successful therapy [208]. In diagnostic clusters, the researcher devises a questionnaire that entails the symptoms or standardized scales. The cluster analysis can then identify groups of patients that present with similar symptoms and simultaneously maximize the difference between the groups. The vector of measurements is arranged into clusters based on similarity where each cluster consists of data points that are more similar to one other than they are to data points of other clusters. Similarities are a set of rules that serve as criteria for grouping or separating items [207].

The two primary clustering methods used in this research were the k-means and the expectation maximization (EM) clustering algorithms. In k-means, a number of k cluster centers are chosen to coincide with k randomly-chosen data points. Then, each pattern of data points is assigned to the closest cluster center. Next, the cluster centers are recomputed. If convergence

criterion is not met, the process is repeated until it is met. Typical convergence criteria are minimal reassignment of patterns to new cluster centers, or minimal decrease in squared error. The EM algorithm is a general purpose maximum likelihood algorithm where it begins with an initial estimate of the parameter vector and iteratively rescores the patterns against the mixture density produced by the parameter vector. The rescored patterns are then used to update the parameter estimates. In a clustering context, the scores of the patterns can be viewed as hints at the class of the pattern. The scores essentially measure the likelihood of being drawn from particular components of the mixture. The patterns, placed in a particular component by their scores, would therefore be viewed as belonging to the same cluster [207].

5.1.3 Clinical Association Rules

Discovery of association rules is an important component of data mining. Association rule learning, also known as affinity analysis, can find patterns in data which reveal combinations of events that occur at the same time. Association rules have wide applicability and have been useful in many areas of nuclear science, pharmacoepidemiology, immunology, bioinformatics, and healthcare [219]. In association mining, the emphasis is almost always on large lifts or positive associations. The two main applications of association mining are market basket analysis and finding prediction rules. In market basket analysis, the dataset consists of a collection of sets 'baskets' and are used to find elements that frequently co-occur together in these sets. Recent studies have shown that there are various algorithms for finding association rules, but one of the best known is the apriori algorithm [219].

5.2 Data Mining Background

Data mining techniques like classification, association, clustering can be applied to healthcare datasets to analyze data to improve diagnosis, health policy-making, early detection of disease outbreaks and preventing the occurrence of various diseases. The techniques provide healthcare authorities an additional source of knowledge for effective decision-making [220]. In order to implement a data mining use case, concepts extracted using a text-mining pipeline supported with a robust terminology seek to answer: *Can co-occurrences of PTSD concepts be exploited in order to gain insight into the data points?*

5.2.1 Taxonomic Support

Desikan et al. [347] provides an introduction to healthcare management and an overview of how data mining helps in analysis of data collected. The authors discuss current trends and prominent models for detection of various diseases. Various types of data used in hospitals like HL7, EHR, EMR, ENR etc. are fully deliberated drawing attention towards the new challenges faced by data mining to aid in healthcare management. Data mining helps in the detection of fraud and abuse, healthcare resource management, diagnosis and treatment of various diseases [220].

Taxonomic arrangement of concepts specifically for PTSD diagnosis is hierarchically organized by criterion in the DSM-V (Diagnostic and Statistical Manual of Mental Disorders, 5^{th}). The diagnostic criterion for clustering analysis in this research utilizes the DSM consisting of Criterion A - Stressor. Criterion G - Functional significance, and four symptom clusters: Criterion B - Intrusion, Criterion C - Avoidance, Criterion D - Negative alterations in cognitions and mood, and Criterion E - Alterations in arousal and reactivity.

5.2.2 Cluster Analysis

This research utilizes a method of clustering PTSD terminology based upon the lexical co-occurrence of concepts. Lexical co-occurrence is an important indicator of word association and this has motivated several co-occurrence measures for word association [348-350]. Concept probabilities of co-occurrence, which is one of several well-known notions of term clustering, forms the basis of this research for grouping of concepts. Term clustering is the grouping of similar words, based on their tendency to co-occur in similar contexts [212]. It was introduced by Brown et al. [213] and used in different applications, including named entity tagging, machine translation, and text categorization.

In most term clustering studies, co-occurrences appearing in the same document, in the same sentence or following the same word has been used to estimate term similarity [212]. Prior research has approached problems of clustering words based upon co-occurrence data, and using the acquired word classes to improve the accuracy of syntactic disambiguation [214]. This research utilizes co-occurrence of concepts as features for machine learning similar to the work of Zhang et al. where co-occurrence features of geo-temporal distributions of tags were extracted and represented as vectors for clustering [215]. According to Jain et al., feature selection is the

"process of identifying the most effective subset of the original features to use in clustering" [207]. This similarity-based clustering of word co-occurrence probabilities is thoroughly described in Cardie in 1993, Ng in 1997, and Zavrel et al in 1997 [216-218]. The feature of co-occurrences of concepts implemented in this research permits the evaluation of terminology coverage for data exploration.

5.2.3 Association Rule Mining

Association rule mining finds interesting associations and/or correlated relationships among large sets of data items that occur frequently together in a given dataset. They provide information of this type in the form of if-then statements. These rules are computed from the data and, unlike the if-then rules of logic, association rules are probabilistic in nature. As the number of items can be several tens of thousands, combinatorics is such that all the rules cannot be studied. It is therefore necessary to limit the search for rules to the most important ones. The quality measurements are probabilistic values which limit the combinatorial explosion during the two phases of the algorithm, and allow the sorting of the results [219].

The definitions for the terms utilized in the association rule mining algorithm for gathering and evaluating were shown in **Table 2.4**. Overall, lift summarizes the strength of association between the products on the left and right hand side of the rule. The larger the lift the greater the link between the two data items as defined in the algorithm implemented [219]. An important characteristic of association rule mining is that it divides the problem of mining into sub-problems to do efficient computing [220].

In Nuwangi et al. [221], the association rule mining is used to generate strong association rules by executing apriori algorithm on real time datasets. The work of these authors demonstrated the apriori algorithm on a dataset by finding frequent itemsets and then generating association rules from frequent sets [220]. Ilayaraja et al.'s research [351] proposed a method using apriori algorithm to find how various diseases occur frequently in a particular geographical area during the year 2012. The outcome also revealed several symptoms occurring simultaneously to examine monthly patterns of patients suffering from heart and liver disease [220]. The experimented results such as these can be used by clinicians to arrive at evidence-based decisions concerning frequently occurring diseases. Combining similar approaches to examine diseases and disorders with other data mining techniques improve results and their

respective interpretation. The research proposed by Patil et al. [352] presents a method which makes use of application of association and classification techniques on numeric data to find whether a patient is likely to be affected by diabetes or not. [220]. Nuwangi et al. [221] researched the global prevalence of diabetes mellitus with the apriori algorithm in Weka. Association rule mining found multiple complications of diabetes in Thailand considering gender, age and occupation factors. This research produced novel, previously inaccessible to the researchers prior to performing the affinity analysis [220]. Witten et al. developed a method using the apriori algorithm and implemented it on a large medical dataset using the proposed technique [202] in Weka. Lekha et al. presented a similar method [223] to generate association rules on numeric data using classification techniques and apriori on a diabetes dataset [220].

5.3 Description of Aim 3

Aim 3. Utilize extracted PTSD signs and symptomatology concepts to perform clustering and affinity analysis for pattern recognition mining, compare results using PTSDO and prior terminological resources.

5.3.1 Conceptual Model

Figure 5.1 describes the conceptual model of aim 3. A text mining pipeline supported by two terminologies are implemented for concept identification of PTSD symptoms. With the extracted concepts from each resource, a cluster analysis and an association rule mining is performed.

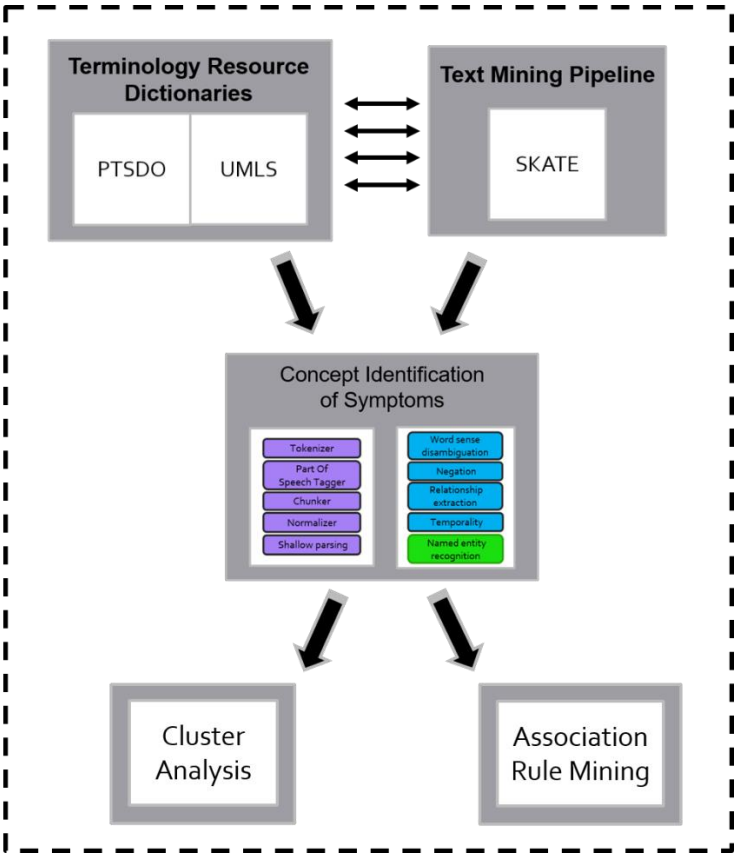


Figure 5.1 Conceptual model of aim 3

5.3.2 Research Questions and Hypotheses

RQ4. *Do PTSD concepts group consistently between clusters in accordance with their hierarchical representations from DSM diagnostic criterion?*

Study 1: A cluster analysis is conducted in order to categorize the co-occurrence of concepts that group together and identify the concepts within each cluster with the highest probability of occurrence.

Study 2: Each concept's groupings are analyzed according to their corresponding DSM diagnostic criterion to explore its respective dispersion for each cluster.

Study 3: The formed clusters to include the concept co-occurrences supported with the PTSD terminology system is compared to the clusters and concept co-occurrences extracted using the UMLS as a base comparison.

Study 4: The data points in each cluster is categorized according to its DSM diagnostic criterion to include the concept co-occurrences and the uniformity of the arrangements supported with PTSDO is compared to those supported with the UMLS. The clusters developed from each terminology source is expected to disperse concepts from each of the DSM diagnostic criteria.

RQ5. *Can association rule mining produce insights into PTSD symptomatology to support hypothesis generation?*

Study 1: Association rule mining is conducted to in order to identify concepts with high degrees of association over a set threshold.

Study 2: The metrics of exact associations rules produced by extraction supported with the PTSDO and the UMLS are compared.

Study 3: Adjustments to minimum threshold is adjusted in order to examine effects on the type and strength of corresponding associations obtained.

Hypothesis 1 – The most commonly identified symptoms among PTSD patients will make up the co-occurrences of formed clusters with the highest probabilities

Hypothesis 2 – There will be a difference in the dispersion of concepts in the clustering supported with the PTSDO compared to the UMLS

Hypothesis 3 – The clusters formed will produce concepts that group uniformly with the DSM diagnostic criterion

Hypothesis 4 – The concepts extracted using the UMLS will produce a larger number of threshold met associations than the concepts extracted using the PTSDO

Hypothesis 5 - There will be no difference in the association metrics produced with the PTSDO and those from the UMLS

5.4 Methods

5.4.1 Dataset and Concept Identification Database

PubMed is a free resource that is developed and maintained by the National Center for Biotechnology Information (NCBI), at the U.S. National Library of Medicine (NLM), located at the National Institutes of Health (NIH). PubMed comprises over 25 million citations for biomedical literature from MEDLINE, life science journals, and online books. PubMed citations and abstracts include the fields of biomedicine and health, covering portions of the life sciences, behavioral sciences, chemical sciences, and bioengineering [21]. The phrase "PTSD Case Reports" entered into the search box and retrieved a total of 1748 citations used for dataset in this research. The filters applied were PubMed Commons and abstract availability reducing the dataset to 1225 citations available for analyses.

The concept identification system used to build a database for analysis is built an adapted version of clinical Text Analysis and Knowledge Extraction System (cTAKES) [200], released open-source at http://www.ohnlp.org. cTAKES is a modular system that builds on existing open-source technologies including the Unstructured Information Management Architecture (UIMA) framework and OpenNLP natural language processing toolkit. This adapted system, referred to as the System for Knowledge Acquisition & Term Extraction (SKATE), implements the pipelined components of cTAKES to provide both rule-based and dictionary based concept recognition. SKATE updates several processes of cTAKES to include a pre-processing regular expression stemming engine.

Major sources of clinical domain concepts are terms in ontologies and controlled terminologies gathered by the Unified Medical Language System (UMLS), which is provided a base comparison for data mining technique analysis to the developed PTSD terminology system. The terminology system, referred to as PTSDO, is modelled to include symptom and treatment vocabulary explicit to the domain of PTSD. The lexicon has been vetted by domain experts and

its coverage has been validated with text mining pipeline entity extraction. Utilizing SKATE to extract concept occurrences attributed to each document, algorithms performed on a database of concepts collected from UMLS support are compared to a database of concepts collected from PTSDO support.

5.4.2 Data Mining Software

The analysis is performed using the Data Mining Client for Excel with SQL Server business intelligence platform [231]. The Data Mining Client makes use of the Microsoft Clustering Algorithm which provides two methods for creating clusters and assigning data points to the clusters. The first, the k-means algorithm, is a hard clustering method. This means that a data point can belong to only one cluster, and that a single probability is calculated for the membership of each data point in that cluster. The second method, the expectation maximization (EM) method, is a soft clustering method. This means that a data point always belongs to multiple clusters, and that a probability is calculated for each combination of data point and cluster [232-233]. Also, several pre-processing tasks were implemented with the Natural Language Toolkit (NLTK) to improve entity extraction [173]. NLTK implemented a regular expression module, a module for stemming, and a tool for data frame formatting. Scikit-learn (http://scikit-learn.org) is executed in order to visualize the cluster analysis. To supplement the Data Mining Client, scikit-learn is an open source machine learning library that produces improved display of co-occurrence distinctness and cluster dissimilarity. arulesViz is an R-extension package that implements several known and novel visualization techniques to explore association rules. Searchable with a unified interactive interface arulesViz creates scatter plots, matrix-based visualizations, graph-based visualizations, and parallel coordinates plots.

5.4.3 Cluster Algorithm Generation

The Microsoft Clustering algorithm is a segmentation algorithm provided by Analysis Services for the Data Mining Client for Excel [231]. The algorithm uses iterative techniques to group cases in a dataset into clusters that contain similar characteristics. The clustering algorithm trains the model strictly from the relationships that exist in the data and from the clusters that the algorithm identifies. It first identifies relationships in a dataset and generates a series of clusters based on those relationships. The clusters group points on the graph and illustrate the

relationships that the algorithm identifies. After first defining the clusters, the algorithm calculates how well the clusters represent groupings of the points, and then tries to redefine the groupings to create clusters that better represent the data. The algorithm iterates through this process until it cannot improve the results more by redefining the clusters. The algorithm can be customized by selecting a specifying a clustering technique, limiting the maximum number of clusters, or changing the amount of support required to create a cluster [232-233, 353].

This research utilized term clustering based upon the occurrence of a concept appearing in the same PubMed case report. This similarity-based clustering of word co-occurrence probabilities utilized concept frequency or concurrence within each PubMed citation. The co-occurrence is used as an indicator of semantic proximity in order to identify the interdependency between PTSD concepts. For similarity metrics, the co-occurrence of concepts is computed by aggregating the number terms that are labeled with their respective concepts. Since the purpose of this modeling task is to estimate the probabilities of co-occurrences, the same co-occurrence statistics are the basis for both the similarity measure and the model predictions [354].

The transactions are accumulated as shown **Table 5.1** which is a subset of the extracted concepts acquired from SKATE supported with PTSDO. For generating the co-occurrence vectors, word counts are calculated from these transactions [355].

PMID	Criterion	Symptom
122472	avoidance	posttraumatic amnesia
549694	reexperiencing	trauma trigger
596752	avoidance	anhedonia
596752	reexperiencing	dream anxiety
879378	arousal	aggressive
879378	avoidance	self-harm
1166290	avoidance	self-blame
1166290	functional impairment	blackout
1358395	trauma	trauma exposure
1442647	reexperiencing	hallucination
1487389	reexperiencing	flashback
1568860	arousal	anger

Table 5.1. Transaction extracted concepts via SKATE supported with PTSDO.

Co-occurrence of words is utilized as the primary means of quantifying semantic relations between words. According to the distributional hypothesis [356] semantically similar words

occur in similar contexts, i.e. they co-occur with the same other words. Therefore, using the immediate co-occurrence of two terms as a measure for their semantic similarity is a way to compare the co-occurrences of the terms with all other terms [357].

For applying the clustering of co-occurrence data, two algorithms were implemented: 1) k-means, and 2) expectation maximization (EM) clustering algorithm. K-means clustering is a well-known method of assigning cluster membership by minimizing the differences among items in a cluster while maximizing the distance between clusters. The means in k-means refers to the centroid of the cluster, which is a data point that is chosen arbitrarily and then refined iteratively until it represents the true mean of all data points in the cluster. The idea in k-nearest neighbor methods is to identify *k* observations in the training dataset that are similar to a new concept to classify. The *k* refers to an arbitrary number of points that are used to seed the clustering process. The k-means algorithm calculates the squared Euclidean distances between data records in a cluster and the vector that represents the cluster mean, and converges on a final set of *k* clusters when that sum reaches its minimum value. Next, similar neighboring concepts to classify the new concept into a cluster, assigning the new concept to the predominant group among these neighbors. The k-means algorithm assigns each data point to exactly one cluster, and does not allow for uncertainty in membership. Membership in a cluster is expressed as a distance from the centroid. Typically, the k-means algorithm is used for creating clusters of continuous attributes, where calculating distance to a mean is straightforward [358-359]. However, the Microsoft implementation adapts the k-means method to cluster discrete attributes, by using probabilities [232]. For discrete attributes, the distance of a data point from a particular cluster is calculated as follows: **1 – P (data point, cluster)** [233].

In EM clustering, the algorithm iteratively refines an initial cluster model to fit the data and determines the probability that a data point exists in a cluster. The algorithm ends the process when the probabilistic model fits the data [231]. The function used to determine the fit is the log-likelihood of the data given the model. If empty clusters are generated during the process, or if the membership of one or more of the clusters falls below a given threshold, the clusters with low populations are reseeded at new points and the EM algorithm is rerun. The results of the EM clustering method are probabilistic. This means that every data point belongs to all clusters, but each assignment of a data point to a cluster has a different probability. Because the method allows for clusters to overlap, the sum of items in all the clusters may exceed the total items in

the training set. In the mining model results, scores that indicate support are adjusted to account for this multiple cluster membership [232, 360-361].

Determining the optimal numbers of clusters for particular datasets is an open problem in the clustering/unsupervised learning research community. There is no definition of the correct amount of clusters to utilize in the modelling nor a principled statistical method to determine the preeminent number of clusters in a data set. The heuristics utilized are often rules of thumb that can have mixed results of best fit for the data. Higher numbers of clusters provide smoothing that reduces the risk of overfitting due to noise in the training data. Generally speaking, if the numbers of clusters is too low, it risks fitting to the noise in the data. However, if cluster count is too high, it misses out on the method's ability to capture the local structure in the data. This research defined the optimal numbers of clusters as the count that achieved superlative classification performance [360-361].

The Microsoft Clustering Algorithm uses a proprietary set of heuristics to best determine the number of clusters to build. This research utilized two methods for exploring optimal numbers of clusters in order to explore beyond the black box heuristic recommendation of the Microsoft Clustering Algorithm [233]. The first is cross-validation technique in order to analyze the number of clusters. In this process, the data is partitioned into v parts. Each of the parts is then set aside at turn as a test set, a clustering model computed on the other v-1 training sets, and the value of the sum of the squared distances to the centroids for k-means is calculated for the test set. These v values are calculated and averaged for each alternative number of clusters, and the cluster number selected that minimizes the test set errors. The second technique, in order to analyze the number of clusters employed, is the elbow method. This method looks at the percentage of variance explained as a function of the number of clusters. The number of clusters is chosen so that adding another cluster does not much improve modeling of the data. More precisely, if one plots the percentage of variance explained by the clusters against the number of clusters, the first clusters will add much, but at some point the marginal gain will drop, giving an angle in the graph. The number of clusters is chosen at this point, hence the elbow criterion [232, 360-361]. In order to display a two-dimensional visualization of the cluster analysis, a Python module package called scikit-learn is implemented. Scikit-learn overcomes the template non-modifiable code extensions by binding compiled libraries [226]. The visualization permits the display of cluster distinctness and the observation of overlap. While the measure of similarity

can be visualized by thickness of connecting lines if any for similarities that meet a threshold for display.

5.4.4 Association Rule Algorithm Generation

The association rules are obtained using the Data Mining Client for Excel making use of the Microsoft Association Algorithm. The algorithm is provided by Analysis Services that is useful for recommendation engines. Association models are built on datasets that contain identifiers both for individual cases and for the items that the cases contain. An association model consists of a series of itemsets and the rules that describe how those items are grouped together within the cases. The rules that the algorithm identifies is used to predict occurrences based on the items that exist in the transactions [231, 233]. A transaction in this study is considered the document-identified case report. The Microsoft Association Algorithm traverses the corpora of case reports to find items that appear together in a case. The algorithm then groups into itemsets any associated concepts that appear, at a minimum, in the number of cases that are specified by the MINIMUM_SUPPORT parameter. The importance of a rule is calculated by the log likelihood of the right-hand side of the rule, given the left-hand side of the rule. The rule format is: $\{I_1\ I_2\} => \{I_k\}$ [232]. The algorithm utilized is a straightforward implementation of the apriori algorithm. This algorithm is the most well-known association rule learning method because it may have been the first [362] and it is very efficient. Algorithm apriori relies on the following subset principle: Every nonempty subset of a large itemset must itself be a large itemset [363-364]. The tasks involved in locating important associations can be summarized in the two steps listed below:

1. **Generate frequent itemsets** that meet the minimum support threshold recursively from 1-itemsets to higher level itemsets, while pruning candidates

2. **Generate rules** that meet the minimum confidence threshold recursively from 1-itemsets to higher level itemsets, while pruning candidates

In the first step, the frequent itemsets are those occurrences that exceed a predefined threshold in the corpus. The second step generates association rules from these itemsets with the constraints of minimal confidence. These frequent itemsets are represented by I_k, $I_k = \{I_1, I_2, \dots ,$

I_k}, association rules with this itemsets are generated in the following way: the first rule is {I_1, I_2, … , I_{k-1}}⇒ {I_k}, by checking the confidence this rule can be determined as interesting or not. Then other rules are generated by deleting the last items in the antecedent and inserting it to the consequent, further the confidences of the new rules are checked to determine the interestingness of them. Interestingness is relative but for this study, it is determined by the lift algorithm described in Chapter 2. These rule-generating processes iterate until the antecedent becomes empty [363-364].

In the association model, rules are based completely on confidence. Therefore, in an association model, a strong rule, or one that has high confidence, might not necessarily be interesting because it does not provide new information [362]. The apriori algorithm does not analyze patterns, but rather generates and then counts candidate itemsets. An item in this research represents an occurrence of a PTSD concept in the case report [231-232].

Using the R-extension package arulesViz package [365], the set of association rules were visually inspected to assist in the identification rules likely to be most useful. Using this package, the rules by confidence, support and lift shown below in **Figure 5.9** and **5.10** are plotted. This plot illustrates the relationship between the different metrics. It has been shown that the optimal rules are those that lie on what's known as the support-confidence boundary. Essentially, these are the rules that lie on the right hand border of the plot where either support, confidence or both are maximized.

5.5 Results

The cluster analysis and association rule learning using extracted concepts acquired from SKATE supported with both PTSDO and UMLS are presented. The results of the analyses supported with PTSDO and a baseline supported with the UMLS are compared. The Diagnostic and Statistical Manual of Mental Disorders (DSM) categories that each concept belongs to is also examined in order to compare the criterion uniformity with each terminology source. Potentially useful associations are discussed as a means of further inquiry and hypothesis generation.

5.5.1 Cluster Analysis

Utilizing the cross-validation technique and the elbow method for determining the optimal number of clusters, a recommendation of the number of clusters is calculated in order to compare

the heuristic number of cluster calculations in the Microsoft Clustering Algorithm. The k-means and the EM algorithms in Microsoft Excel Data Mining plug-in for clustering terms is applied and the results are presented. The UMLS is implemented and used as a baseline of results in order to compare clustering utilizing the PTSDO. After clustering, the applicable concept's DSM criterion is observable for each grouping.

5.5.1.1 K-means Clustering with UMLS

The Microsoft heuristic calculations determined the number of clusters to be six. Both the elbow method and the cross-validation of concepts method concurred with the number of clusters. Displayed in **Figure 5.2**, cluster one in the k-means analysis supported with the UMLS contains a total of 12 concepts. The most prominent concept is C0001807: aggressive behavior with a confidence of occurring at 38%. Other prominent concepts above a 5% probability includes C0344315: depression, C0178417: anhedonia, C0517894: relationship difficulty, and C0424366: self-harm. Cluster two in the k-means analysis consists of 10 total concepts. C3714660: trauma exposure concept was the most frequent with a confidence of occurring at 45%. Prominent concepts in cluster two above a 5% probability includes C0004448: awareness, C0575090: balance impairments, and C0033213: issues. Cluster three consists of a total of 6 concepts. The most frequent concept is C0871189: psychotic behavior with a confidence of occurring at 19%. Other numerous concepts in cluster three above a 5% probability includes C2919017: cognitive distortion, C0599437: authority difficulty, C0039869: thought, and C0558058: reflecting. Cluster four in the k-means analysis consists of a total of 7 concepts. The most frequent concept is C0003808: arousal reactivity with a confidence of occurring at 15%. Other numerous concepts above a 5% probability includes C2587213: control, C0231303: distress, C0030193: pain, C0917801: insomnia, and C0338831: mania in the fourth cluster. 5 concepts appear in the fifth cluster. The most numerous concept is C0730557: emotional recollection with a confidence of occurring at 18%. Other frequent concepts in the fifth cluster above a 5% probability includes C0013117: dream, C0030318: panic, C0236720: flashback, C0561837: intrusive memory, C0018524: hallucination, and C0233488: despair. Lastly, cluster six in the k-means analysis contains a total of 7 concepts. The most numerous concept is C0917801: insomnia with a confidence of occurring at 26%. Other frequent concepts in the sixth

cluster above a 5% probability included C0030318: anxious reaction, C0002957: anger, C0040678: transference, and C1446377: mental health problems.

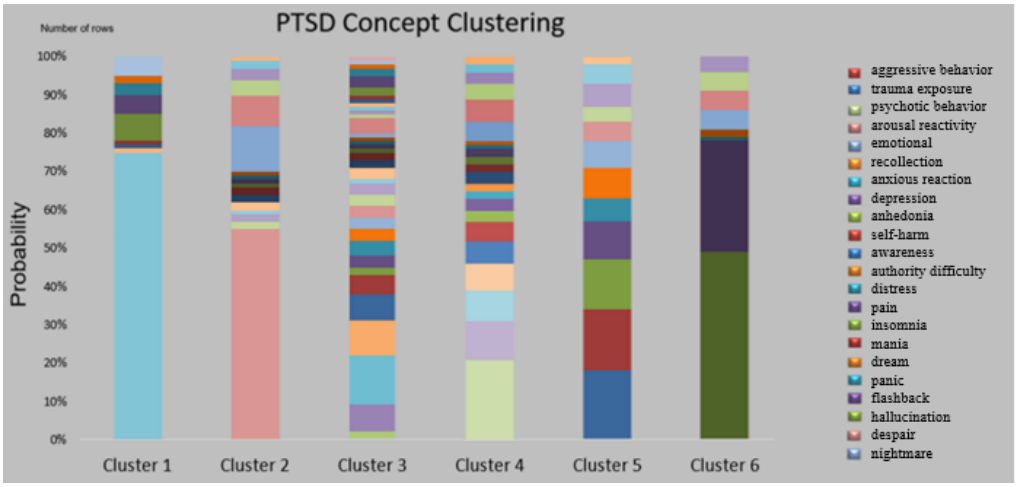


Figure 5.2. K-means PTSD concept clustering with UMLS

The clusters are implemented in scikit-learn in order to display visual clustering of the occurrence of concepts [226]. Although the cluster vectors are three-dimensional data points, a two-dimensional format is visualized to display similarity between clusters. **Figures 5.3a** and **5.2b** display the data points in a two-dimensional format which do not necessary show the similarity of the clusters. However, an overlay of connected-lines produced in the Data Mining Client are used to display similarity between clusters scaled to the ratio of likeness between the groupings of data points. The darker the lines indicate greater similarity with the non-existence of lines indicate little to no similarity between clusters. For instance, **Figure 5.3a** displays a dark line between cluster two and cluster five representing a large amount of similarity between the clusters. There is no line between cluster one and cluster two representing great dissimilarity between the two clusters. This figure displays a slightly shaded line between cluster one and cluster four representing a small amount of similarity between the clusters. The co-occurrence concepts obtained in the clustering analysis are also transformed according to their respective DSM criterion. This display is used to answer the question: *Do PTSD concepts group consistently between clusters in accordance with their hierarchical representations from DSM diagnostic criterion?*

A total of 33 concepts that are clustered do not belong to a DSM category. Non-DSM criteria concepts for cluster one includes C0012725: displacement, C0233677: nihilistic despair, C2987481: strain, C3714756: intellectual disabilities, and C0041657: unconscious. Cluster two includes non-DSM categories of C0575090: balance impairments, and C0033213: issues. Additionally, C0558058: reflecting, C0871189: thought, C0700327: irrepressible memories, C0026773: dissociative identity disorder, C0233522: inappropriate behavior, C0231242: complicated, C2584308: intimate relationship adjustment, C3263722: injury, and C0575090: balance impairments are included in cluster three. Non-DSM criteria concepts for cluster four include C1546466: problems, C000967: conflict, C0004930: behavioral disorder, C0012833: dizziness, C0442801: exaggerate, C0233622: ritual compulsion, C0525045: mood disorder, and C0040822: tremor. Cluster five includes non-DSM categories of C0233488: despair, C2987481: strain, and C0004448: awareness. Lastly, C0040678: transference, and C1446377: mental health problems are included in cluster six.

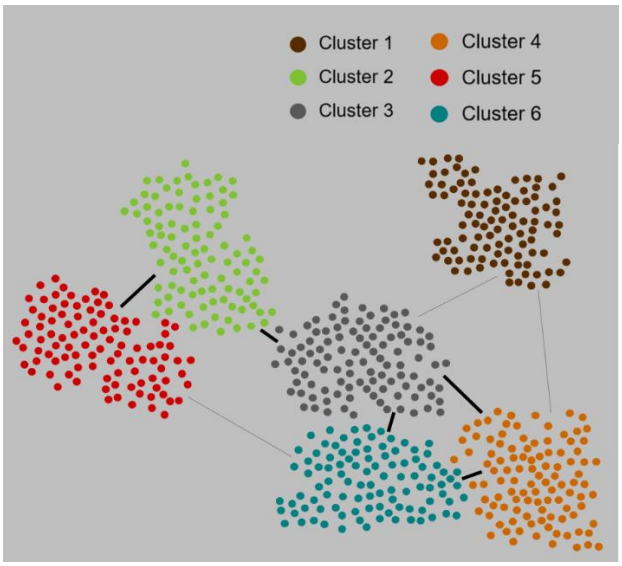


Figure 5.3a. scikit clustering supported with UMLS

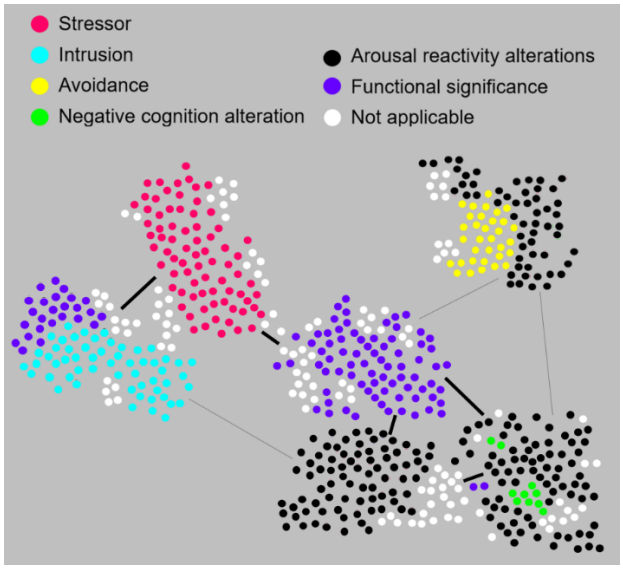


Figure 5.3b. DSM criterion of features

Within the clustering supported with the UMLS, each cluster is primarily dominated by one or two DSM diagnostic criterion and each grouping consists of concept co-occurrences that do not fit into a DSM category. For example, cluster two consists of primarily stressor criterion concepts along with only a few non-DSM criterion concepts. Cluster six similarly consists of only one criterion arousal reactivity alterations and the remaining are concepts not contained in

the DSM. Only three of the clusters contains more than two DSM criteria concepts in a single grouping.

The avoidance and negative cognition alteration criterion, and intrusion are considerably underrepresented in clusters with each only being represented in a single cluster. Arousal is represented in half of the clusters. Functional significance concept occurrence is minimally represented in two clusters. The percentages of DSM criterion for the six clusters is highlighted in **Figure 5.4**. For example, cluster one contains 52% of arousal criterion category, 34% avoidance, and 14% of concepts not included in the DSM PTSD criterion. Of interest, cluster four shows 73% in the arousal criterion, 15% non-applicable to DSM, 10% negative alterations, and 2% functional significance. Stressor criteria concept occurrences are well represented as the majority of probability in 3 clusters. The dispersion of concept occurrences in any cluster does not correspond uniformity to the DSM. Each cluster consists of one or a few DSM criteria concepts and the lack of dispersion is apparent.

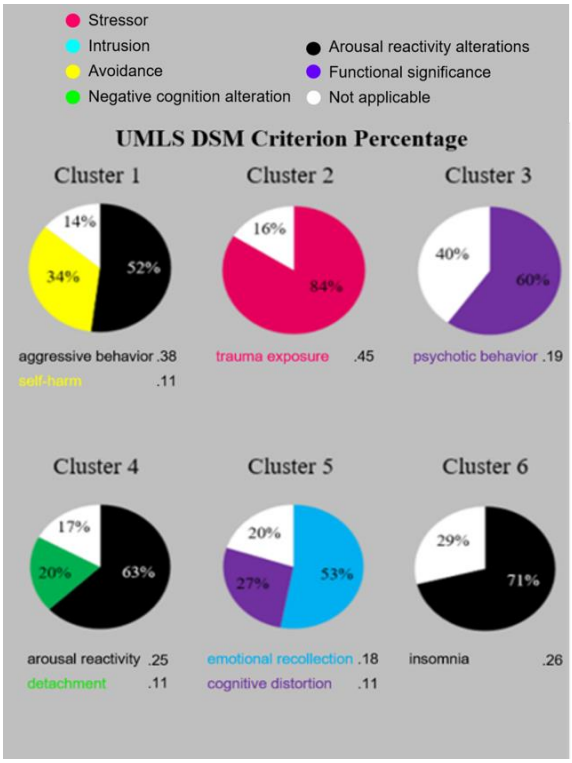


Figure 5.4. UMLS k-means DSM criterion clustering

5.5.1.2 K-means Clustering with PTSDO

Utilizing the cross-validation technique for determining the optimal number of clusters, a recommendation of six clusters is selected, similar to the heuristics utilized in the Microsoft Clustering Algorithm. The elbow method recommends a count of seven clusters. The cross-validation and clustering algorithm recommendation of six is employed for the clustering technique.

K-means clustering of extracted concepts supported with PTSDO results in six clusters shown in **Figure 5.5** with the number of co-occurrence averaging 12 concepts among groupings. Cluster one in the k-means analysis contained a total of 13 concepts. The most prominent concept is P0000010: trauma exposure with a confidence of occurring at 53%. Other prominent concepts above a 5% probability includes C0871693: combat exposure, P0000026: authority difficulty, and C0178417: anhedonia.

Cluster two in the k-means analysis contains 10 concepts. The most frequent concept is C0848237: acute stress with a confidence of occurring at 48%. Concepts above a 5% probability includes P0000474: feel inadequate, C2957419: military combat stress, C1387813: dream anxiety, and C0277785: dysfunctional. A total of 16 concepts makes up the third cluster. P0000280: trauma trigger concept is the most frequent with a confidence of occurring at 28%. Prominent concepts in cluster three above a 5% probability includes P0000543: forget medication, C0871189: psychotic symptom, P0000546: relationship difficulty, and C0231303: distress. Cluster four in the k-means analysis consists of a total of 11 concepts. The most frequent concept is C0038436: post-traumatic stress with a confidence of occurring at 20%. Other numerous concepts above a 5% probability includes C0424366: self-harm, C1821940: flashback, and C0018524: hallucination in the fourth cluster. Distributed in **Figure 5.5**, 14 concepts appear in the fifth cluster. P0000012: arousal reactivity with a confidence of occurring at 22% is the most recurrent concept in this cluster. Above a 5% probability, the concepts P0000550: cognitive distortion, P0000059: acute panic, C0730557: emotional abuse, P0000284: social dysfunction, C0030193: pain, and C1963237: insomnia met this threshold. Lastly, the k-means analysis in the sixth cluster contains a total of 12 concepts. The most frequent concept is C1579931: depression with a confidence of occurring at 18%. Other numerous concepts in cluster six above a 5% probability included C0849912: emotional recollection, P0000536:

employment difficulty, C2587213: control issue, C0038580: substance dependence, and C3263723: traumatic injury.

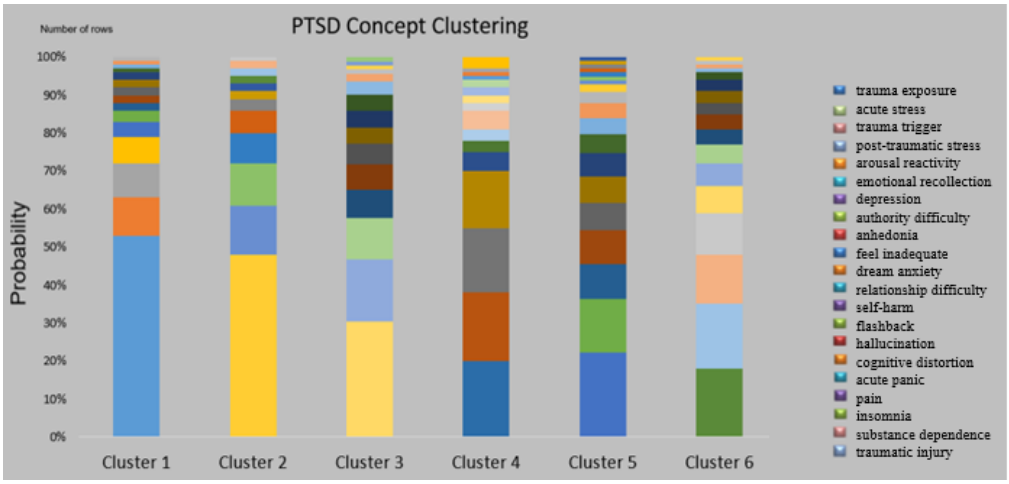
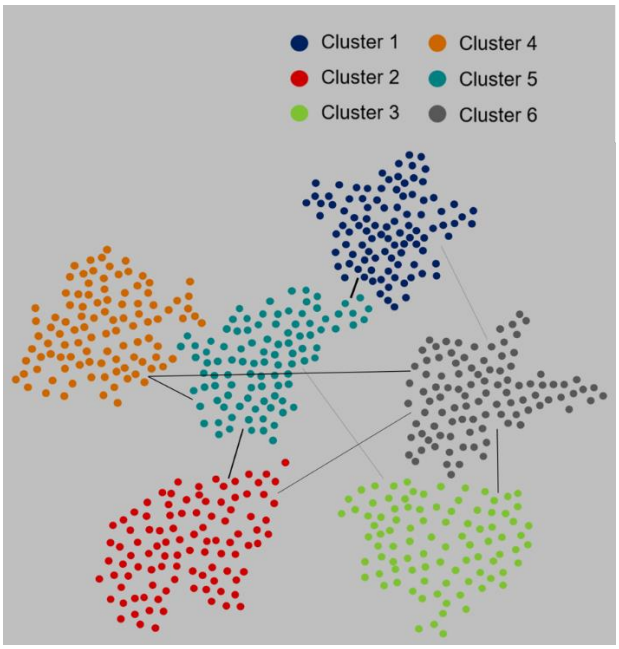


Figure 5.5. K-means concept clustering with PTSDO

Figures 5.6a and **5.6b** display the two-dimensional clustering visualized with scikit-learn [226]. **Figure 5.6a** displays a dark line between cluster one and cluster five representing a large amount of similarity between these clusters. There is lightly shaded line between cluster three and cluster five representing only slight similarity between the two clusters. This minimal amount of similarity also exists between cluster one and cluster six as well.

The co-occurrence concepts obtained in the clustering analysis are also transformed according to their respective DSM criterion in order to examine their hierarchical representations of DSM diagnostic criterion. The clustering determined by the PTSDO appeared to cluster more uniformly according to DSM criteria than the clustering supported with the UMLS. However, there appears to be much similarity between the six clusters making the groupings less distinct than the UMLS.



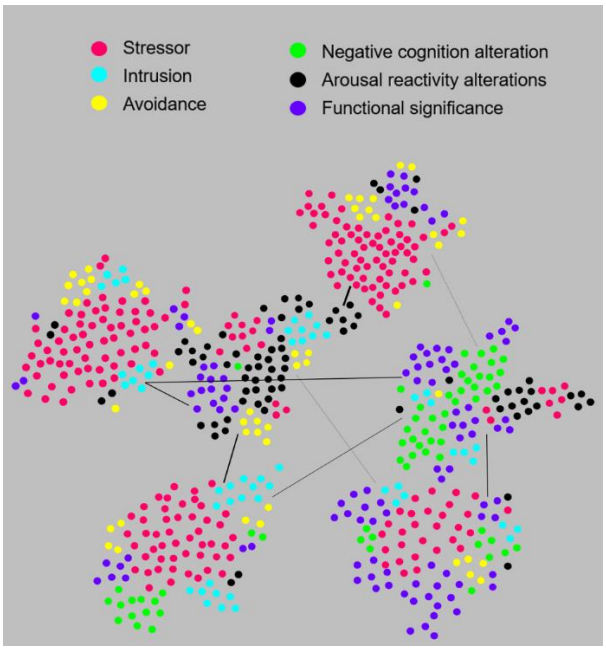


Figure 5.6a. K-Means clustering supported with PTSDO **Figure 5.6b.** DSM criterion of features

The percentages of DSM criterion in clustering supported with PTSDO are shown in **Figure 5.7**. Each cluster consists of concept co-occurrences from all PTSD DSM category of which the most prominent in each criterion is listed in the figure. Each criterion is represented in every grouping. None of the formed clusters are dominated by one or two DSM diagnostic criterion but rather a dispersion. An outlier is cluster one that is made up of a 69% probability of occurrence of a stressor criterion. Of interest, cluster five shows 49% in the arousal criterion, 16% functional significance, 13% stressor, 9% re-experiencing, and 1% negative alterations. Stressor criteria concept occurrences is well represented as the majority of probability in two clusters. Functional significance and re-experiencing is also respectively dispersed among the clusters. The dispersion of PTSD concept occurrences within each cluster display some amount of uniformity corresponding to the DSM.

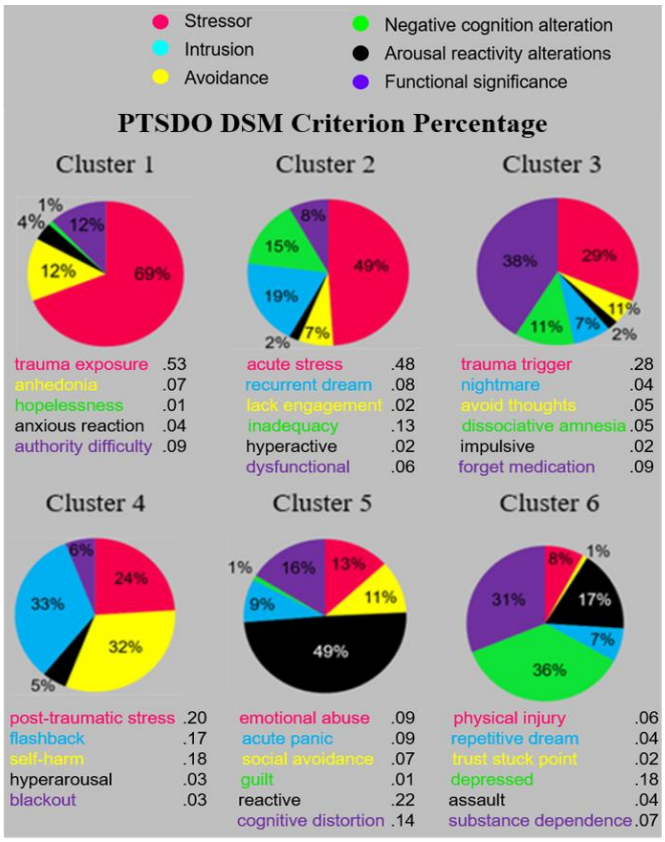


Figure 5.7. PTSDO k-means DSM criterion clustering

5.5.1.3 Expectation–Maximization (EM) Clustering with PTSDO

The Microsoft heuristic calculations of the number of clusters was determined to be four and cross-validation of concepts supported this count to seem to work the best. EM clustering of extracted concepts supported with PTSDO results in four clusters shown in **Figure 5.8** with the number of co-occurrence averaging 17 concepts among groupings.

Cluster one in the EM analysis contained a total of 10 concepts. The most prominent concept is P0000010: trauma exposure with a confidence of occurring at 53%. Other prominent concepts above a 5% probability includes C0871693: combat exposure, P0000026: authority difficulty, and C0178417: anhedonia. A total of 9 concepts makes up the second cluster. Cluster 2 in the EM analysis contained 11 concepts. C0277785: dysfunctional is the most frequent with a confidence of occurring at 18%. Concepts above a 5% probability includes P0000474: feel inadequate, C0424092: withdrawn, C2957419: military combat stress, C1387813: dream anxiety, and C0848237: acute stress. 12 concepts appear in the third cluster. P0000280: trauma trigger with a confidence of occurring at 28% is the most recurrent concept in this cluster. Above a 5%

probability, the concepts C0018524: hallucination, P0000543: forget medication, P0000546: relationship difficulty, and C0231303: distress met this threshold. Lastly, the EM analysis in the fourth cluster contained a total of 10 concepts. The most frequent concept is C0338831: mania with a confidence of occurring at 19%. Other numerous concepts in cluster six above a 5% probability includes C0424366: self-harm, C0344315: depression, C0338831: mania, C0871189: psychotic symptom, C1821940: flashback, and C0038436: post-traumatic stress.

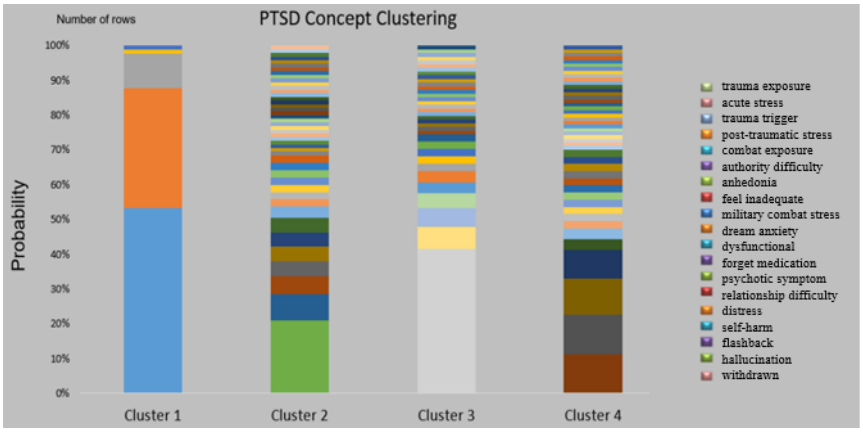


Figure 5.8. EM concept clustering with PTSDO

Two-dimensional clustering is visualized with scikit-learn in **Figures 5.9a** and **5.9b** [226]. Shown in **Figure 5.9a**, there is great similarity between cluster two and cluster four. These two clusters could have been combined into a single cluster, however the heuristic cross-validation recommended the four clusters as opposed to three. There is lightly shaded line between cluster one and cluster three representing only slight similarity between the two clusters. This minimal amount of similarity also exists between cluster three, cluster two and four as well. The co-occurrence concepts obtained in the clustering analysis are also transformed according to their respective DSM criterion in order to examine their hierarchical representations of DSM diagnostic criterion.

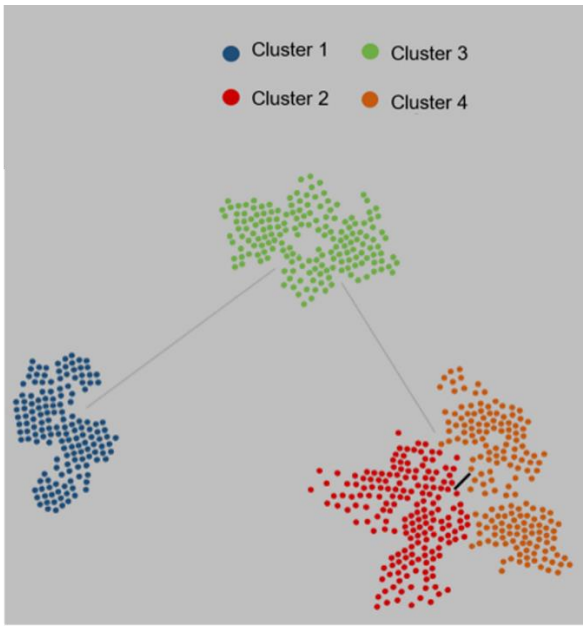


Figure 5.9a. EM clustering supported with PTSDO

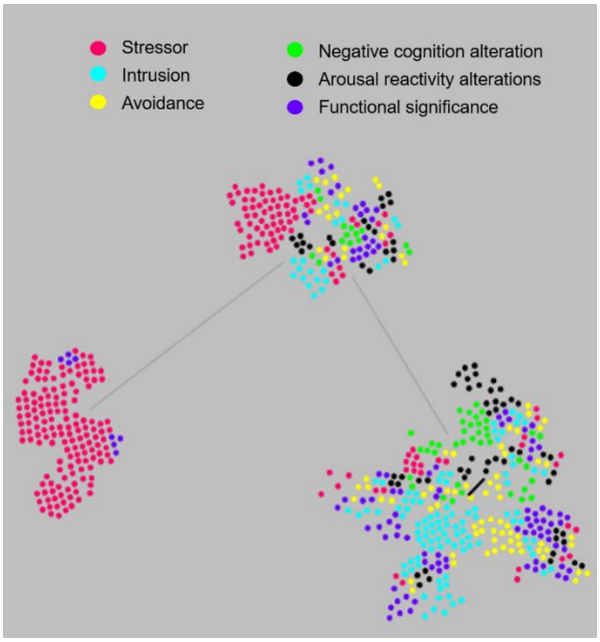


Figure 5.9b. DSM criterion of features

EM clustering with PTSDO creates moderately DSM criterion uniformity among its four clusters with the exception of cluster one. 88% of cluster one is made up of a stressor with the reaming 12% functional significance concepts. Each of the remaining clusters contain concepts from all DSM categories and are fairly evenly dispersed.

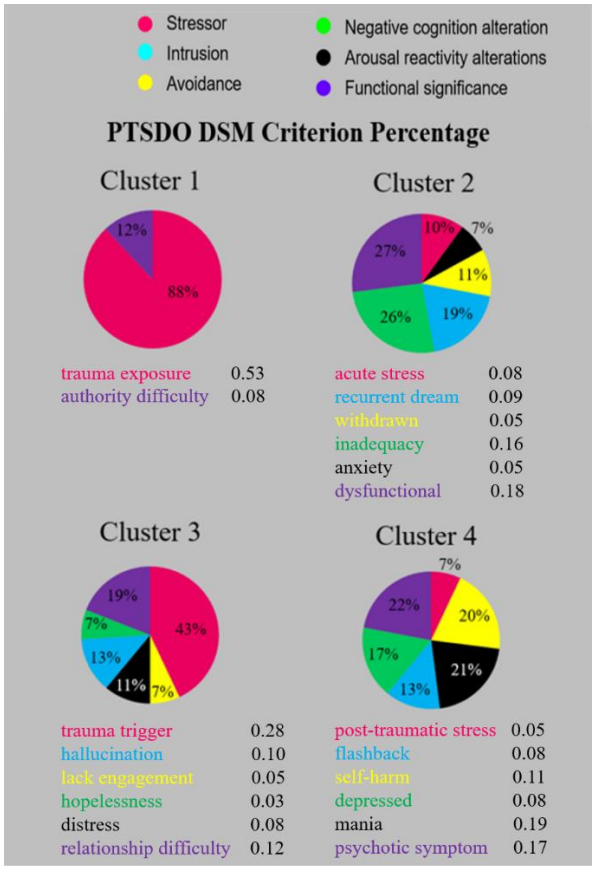


Figure 5.10. EM PTSD concept DSM criterion clustering supported with PTSDO

5.5.2 Association Rule Mining

The apriori algorithm is run on the concept-based transactions with a support threshold of 22, meaning the minimal level of co-occurrences. Additionally, any rules with confidence less than 40% is eliminated. A substantial collection of association rules is attached in **Appendix I**. For each rule, the table indicates the support and confidence. The rules are listed in decreasing order of lift. Shown in **Figure 5.11**, the R-extension package arulesViz package displays a scatterplot of confidence, support and lift metrics from the UMLS. Each data point in the scatterplot represents an occurrence of an association rule. Many of the rules with higher lifts have relatively lower support. Several interesting rules can be identified along the support and confidence border as shown in the plot. Not shown in the figure, arulesViz provides many interactive features for exploring and allowing association rules to be identified with ease.

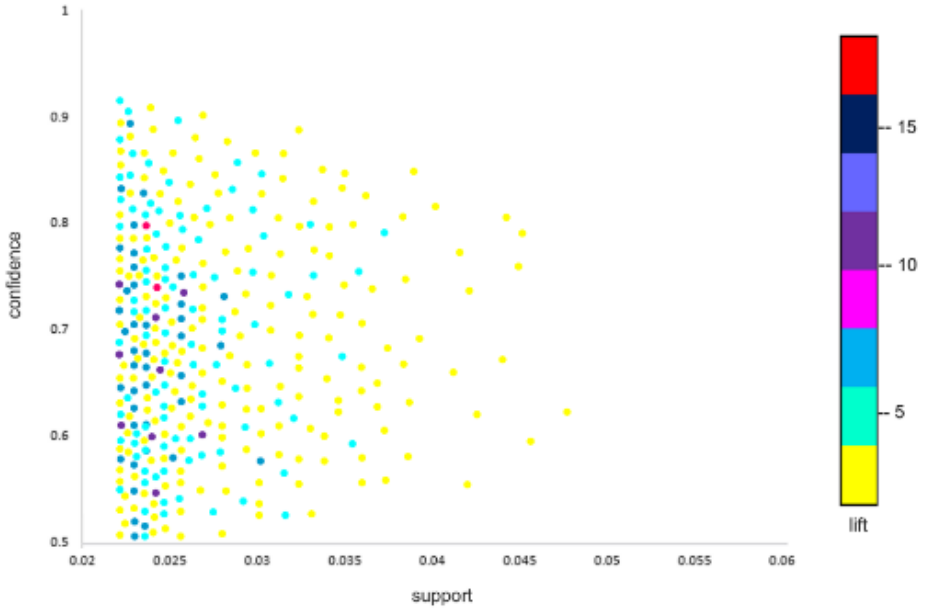


Figure 5.11. UMLS scatter plot of confidence, support and lift metrics

The interactive feature of arulesViz allows the selection of individual rules in the scatter plot and inspecting each individually. Also, sets of rules can be inspected together by selecting a rectangular region of the plot. Searching the plot is efficient with zooming in and out features as well as the ability to filter out rules. As an example, **Table 5.2** identifies rules extracted with relatively good lift and correspond to trauma exposure. These rules are deemed interesting as the etiology of PTSD is such that the disorder is precipitated by a trauma or stressor. Concepts that make up the rule are listed on the left-hand side which correspond to a concept on the right-hand side. Each association rule's respective support, confidence, and lift metrics are attached.

Association Rules						
ID #	LHS		RHS	Support	Confidence	Lift
1	{ self-harm, amnesia }	=>	{trauma exposure}	0.25	0.71	9.47
2	{ memory impairment, fear }	=>	{trauma exposure}	0.26	0.64	9.16
3	{ anger, feel inadequate }	=>	{trauma exposure}	0.23	0.62	8.85
4	{ blackout, acute stress, mania }	=>	{trauma exposure}	0.25	0.61	6.12
5	{ memory impairment, suicidal, self-harm, agoraphobia }	=>	{trauma exposure}	0.23	0.57	6.09
6	{ memory impairment, relationship difficulty, feel inadequate}	=>	{trauma exposure}	0.22	0.52	5.91

Table 5.2. UMLS concepts corresponding to trauma exposure

After consulting the arulesViz scatterplot shown in **Figure 5.12** of associations acquired from extracted concepts supported with PTSDO, rules of interest can be efficiently searched. Associations that have higher lift can be researched using the interactive tool or filters can be set to mine for specific correspondence. Filtering is implemented in order to identify similar associations identified in **Tables 5.2** for comparison of the rules identified in **Table 5.3**, which are described below.

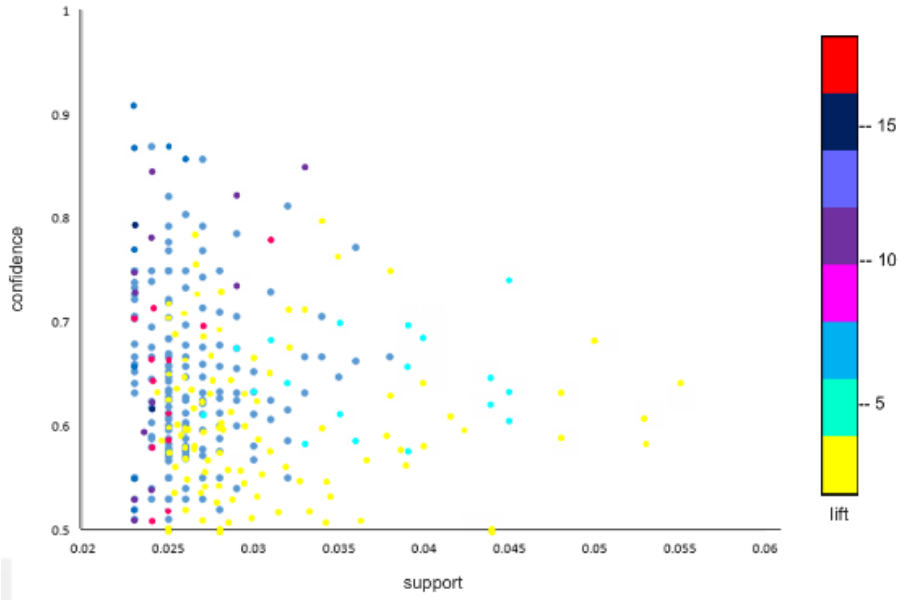


Figure 5.12. PTSDO scatter plot of confidence, support and lift metrics

Association Rules						
ID #	LHS		RHS	Support	Confidence	Lift
1	{ memory impairment, control issue }	=>	{trauma exposure}	0.25	0.66	12.72
2	{ self-harm, cognitive distortion }	=>	{trauma exposure}	0.29	0.61	11.64
3	{ memory impairment, relationship difficulty, academic difficulty, feel inadequate}	=>	{trauma exposure}	0.23	0.59	9.72
4	{ blackout, acute stress, mania }	=>	{trauma exposure}	0.30	0.67	8.92
5	{ memory impairment, control issue, self-harm, cognitive distortion }	=>	{trauma exposure}	0.23	0.58	8.20
6	{ academic difficulty, feel inadequate }	=>	{trauma exposure}	0.23	0.65	6.88

Table 5.3. PTSDO concepts corresponding to trauma exposure

Comparable rules extracted with PTSDO to those from UMLS are shown in **Table 5.2**. Displayed are left-hand side symptom concepts that correspond to the concept of trauma exposure. Rule No. 4 in the PTSDO table is exactly the same as rule No. 4 in the UMLS table. The only difference is the slightly higher metrics for the rule in the PTSDO table. This is likely

due to a higher count of extracted concepts the rule contains due to better coverage. This improved coverage can be identified in the greater number of synonyms in PTSDO for the concepts of C0312422: blackout, C0848237: acute stress, and C0678190: mania. Also very similar, rule No. 2 in the PTSDO table and rule No. 1 in the UMLS table both have the left-hand side concept C0424366: self-harm. The only difference is again slightly higher metrics for the PTSDO rules and the concept of cognitive distortion of which the UMLS has a more detailed type of cognitive distortion. For these PTSDO association rules, the UMLS contained the concepts of C0233794: memory impairment, C0517894: relationship difficulty, C0847487: feel inadequate, C0424366: self-harm. The UMLS does not the concepts of academic difficulty, control issue, and cognitive distortion. Available concepts for post-coordination from the UMLS to produce academic difficulty included C1510747: academia and C1299586: has difficulty doing. UMLS concepts to produce control issue with post-coordination included C1287165: self-control and C0150632: impulse control. Existing concepts for post-coordination from the UMLS to produce cognitive distortion included C1516691: cognitive and C2919017: distortion. Observable from the extractions, there is a greater number of association rules corresponding to trauma exposure acquired from the baseline of UMLS compared to those from the PTSDO. However, there are more association rules with greater lift acquired from the PTSDO.

Table 5.4. lists several interesting association rules from concepts extracted supported with PTSDO. Many of the rules identify symptoms that correspond to a stressor concept. This could be very helpful in identification of the disorder when a full screening of a patient is not obtained or available. This data mining technique displays how only a few number of symptom concepts could be used to identify further inquiry into the status of a patient or trauma victim. This is especially important when considering the large spectrum of symptoms used to acquire diagnosis with the DSM or the lengthy patient checklist applied when screening individuals of concern. Below, several rules of interest are presented along with interesting relationships that are supported by current knowledge and research.

Association Rules						
ID #	LHS		RHS	Support	Confidence	Lift
1	{memory impairment, relationship difficulty, academic difficulty, feel inadequate}	=>	{trauma exposure}	0.23	0.59	9.72
2	{memory impairment, control issue, self-harm, cognitive distortion}	=>	{trauma exposure}	0.23	0.58	8.20
3	{blackout, acute stress, mania}	=>	{trauma exposure}	0.30	0.67	8.92
4	{memory impairment, relationship difficulty}	=>	{trauma exposure}	0.24	0.65	10.08
5	{blackout, acute stress}	=>	{trauma exposure}	0.25	0.67	11.18
6	{blackout}	=>	{trauma exposure}	0.24	0.66	9.84
7	{memory impairment, control issue}	=>	{trauma exposure}	0.25	0.66	12.72
8	{mania, forget medication}	=>	{trauma exposure}	0.25	0.65	8.24
9	{academic difficulty, feel inadequate}	=>	{trauma exposure}	0.23	0.65	6.88
10	{self-harm, cognitive distortion}	=>	{trauma exposure}	0.29	0.61	11.64
11	{pain, cognitive distortion}	=>	{trauma exposure}	0.25	0.60	6.24
12	{self-harm, forget medication}	=>	{trauma exposure}	0.24	0.46	4.20
13	{feel inadequate}	=>	{trauma exposure}	0.24	0.48	4.44
14	{feel inadequate}	=>	{blackout}	0.25	0.51	7.28
15	{self-harm, hallucination}	=>	{post-traumatic stress}	0.27	0.66	11.45
16	{self-harm, withdrawn, blackout}	=>	{post-traumatic stress}	0.26	0.68	11.28
17	{self-harm, difficulty communicating}	=>	{post-traumatic stress}	0.25	0.67	10.84
18	{self-harm, blackout, flashback}	=>	{post-traumatic stress}	0.26	0.71	10.08
19	{self-harm, agoraphobia}	=>	{post-traumatic stress}	0.23	0.64	9.94
20	{self-harm, blackout}	=>	{post-traumatic stress}	0.24	0.58	8.93
21	{self-harm}	=>	{post-traumatic stress}	0.23	0.52	7.65
22	{blackout, hyperarousal}	=>	{post-traumatic stress}	0.30	0.51	6.45
23	{blackout}	=>	{post-traumatic stress}	0.25	0.64	2.08
24	{withdrawn, flashback, difficulty communicating}	=>	{post-traumatic stress}	0.23	0.81	5.39
25	{withdrawn, difficulty communicating, blackout}	=>	{post-traumatic stress}	0.26	0.78	3.39
26	{withdrawn, hyperarousal, agoraphobia}	=>	{post-traumatic stress}	0.26	0.64	3.11
27	{withdrawn, self-harm, blackout}	=>	{post-traumatic stress}	0.25	0.59	2.75
28	{withdrawn, self-harm}	=>	{post-traumatic stress}	0.30	0.41	2.36
29	{social dysfunction, insomnia, blunted affect}	=>	{headache}	0.24	0.62	6.21
30	{social dysfunction, blunted affect}	=>	{headache}	0.23	0.54	5.54
31	{blunted affect, mania, temper}	=>	{headache}	0.25	0.52	5.33
32	{social dysfunction, blunted affect, temper, fear}	=>	{pain}	0.29	0.48	8.84
33	{social dysfunction, blunted affect}	=>	{pain}	0.27	0.52	6.79
34	{blunted affect}	=>	{pain}	0.25	0.63	4.37
35	{anhedonia, feeling isolated, difficulty coping, anger}	=>	{trauma exposure}	0.28	0.52	11.48
36	{anxious reaction, feeling isolated, difficulty coping}	=>	{trauma exposure}	0.27	0.44	11.13
37	{feeling isolated, suicidal, difficulty coping, anger}	=>	{trauma exposure}	0.27	0.46	9.48
38	{feeling isolated, difficulty coping, anger}	=>	{trauma exposure}	0.26	0.48	8.89
39	{anxious reaction, suicidal, anger}	=>	{trauma exposure}	0.29	0.61	8.87
40	{anxious reaction, suicidal, hopelessness}	=>	{trauma exposure}	0.28	0.59	8.45
41	{anxious reaction, anger}	=>	{trauma exposure}	0.30	0.68	8.12
42	{anger, feeling isolated, difficulty coping}	=>	{anxious reaction}	0.27	0.57	4.19
43	{hopelessness, feeling isolated, difficulty coping}	=>	{anxious reaction}	0.26	0.54	4.01
44	{feeling isolated, difficulty coping, anger}	=>	{suicidal}	0.25	0.53	3.45
45	{feeling isolated, difficulty coping}	=>	{suicidal}	0.25	0.63	3.23
46	{vulnerable, feel inadequate, dysfunctional}	=>	{acute stress}	0.25	0.52	4.48
47	{vulnerable, dysfunctional, feel numb, dream anxiety}	=>	{acute stress}	0.23	0.48	3.74
48	{vulnerable, feel inadequate, dream anxiety}	=>	{acute stress}	0.24	0.57	3.17
49	{dysfunctional, dream anxiety, lack engagement}	=>	{acute stress}	0.25	0.51	3.03
50	{feel inadequate}	=>	{acute stress}	0.24	0.46	2.14
51	{vulnerable, lack engagement, feel numb}	=>	{feel inadequate}	0.24	0.48	2.11
52	{vulnerable, feel numb}	=>	{feel inadequate}	0.23	0.48	2.06
53	{amnesia, helplessness, dissociative, memory impairment}	=>	{trauma trigger}	0.24	0.58	8.57
54	{amnesia, distress, dissociative, memory impairment}	=>	{trauma trigger}	0.24	0.49	8.18
55	{amnesia, nightmare, dissociative, memory impairment}	=>	{trauma trigger}	0.23	0.47	7.63
56	{amnesia, nightmare}	=>	{trauma trigger}	0.26	0.56	4.42
57	{helplessness, nightmare}	=>	{trauma trigger}	0.25	0.53	4.13
58	{distress, helplessness}	=>	{amnesia}	0.24	0.48	3.87
59	{distress, helplessness, nightmare, memory impairment}	=>	{amnesia}	0.24	0.48	6.42
60	{distress, helplessness, memory impairment}	=>	{amnesia}	0.25	0.42	5.21
61	{amnesia, distress, memory impairment}	=>	{dissociative thought}	0.24	0.41	2.45
62	{amnesia, helplessness, distress}	=>	{memory impairment}	0.23	0.49	2.49
63	{amnesia, helplessness, nightmare}	=>	{memory impairment}	0.23	0.46	2.42

Table 5.4. Association rules obtained from PTSDO

The threshold for acquired rules for analysis is set to extracted concepts and association rules above 40% probability of occurrence. With the PTSDO, the total number of association rules above 40% probability of occurrence collected was 218 compared to 293 association rules gathered from the baseline implementation of extracting concepts from the UMLS. However, 74 association rules from the UMLS baseline contained concepts that were not relevant to exploring information about PTSD or improving its understanding.

As seen in **Table 5.4**, several rules are made up of the concept of C0424366: self-harm which correspond to the stressors of trauma exposure and post-traumatic stress. Its high confidence value for Rule No.21 is somewhat perplexing. The single concept corresponds to trauma exposure with support of 0.23, confidence of 0.52, and lift of 7.65. A reasonable correspondence exists between self-harm with blackout in Rule No. 20 but the concepts seem to have no relationship with one another. One plausible explanation for such a rule is that the concepts do correspond to one another but at a lower threshold.

Rules No. 51 and 52 demonstrate a logical relationship between avoidance symptoms corresponding to feelings of inadequacy. Rules No. 35-38 represent logical relationships of several DSM clusters corresponding to the stressor of trauma exposure. Rule No. 57 shows a relationship between the two concepts of helplessness and nightmare together corresponding to a trauma trigger. The concept of helplessness occurs again in Rules 62 and 63 displaying correspondence with the two avoidance concepts of amnesia and memory impairment. This is a logical rule considering helplessness is a dissociative sign or symptom sometimes considered to be a form of a coping mechanism. One of the determining factors of a trauma trigger seems to be the feeling of extreme helplessness at the time the trauma occurred. Those that do not experience a helpless feeling are often able to process the event normally. Rule No. 32 portrays the defense mechanism of avoidance that continues in a PTSD patient and has been shown to produce a paradoxical result of long-term experiences of general pain. More specifically, Rules No. 29, 30, and 31 highlight the relationship of avoidance and numbing concepts to the specific pain of headaches. Additionally, there are rules displaying out of body experiences corresponding to instances of pain. The out of body experiences may also explain why at other times a patient's body feels disconnected or weak down one side. Rules No. 39-41 show relationships between anxious reactions and trauma exposure. Anxiety is one of the most common symptoms associated with posttraumatic stress that manifests itself with other symptoms. Rules No. 39 and

41 include signs and symptoms of anxious reaction and suicidal ideation along with anger. This is logical because anger can be predictive of PTSD severity and experiences with disassociation can impair functions in a patient.

Interestingly, Rules No. 57-58 and 62-63 portray correspondence between helplessness, avoidance concepts, and dissociative signs or symptoms. According to Freud's theory of psychological defenses, emotional trauma coupled with helplessness can lead the mind to protect itself from a sudden influx of extreme stimulus by diverting it away from the normal flow of thought and feeling [369]. When someone is overtaken unaware by a traumatic event, however, they can't rely on their usual resources to process or contain the flood of fear, anxiety, pain, etc. aroused by the event. To protect itself from this overwhelming experience, the mind erects a barrier against the memory, segregating it from other memories and emotions in a process of defense against it. This is shown in rules No. 32-33 with emotions having strong metrics that correspond to pain. The defense fails as the memory can't be entirely excluded but continues to exert an extremely powerful effect. The paradoxical result of this effort to ward off an overwhelming experience is to give it a lasting power to cause pain. Rules No. 36 and No. 39-41 show several arousal and anxiety association rules with very respectable metrics. Fear and anxiety can intensify dissociative symptoms. Extracting more knowledge, such as these identified rules, can aid in reducing the intensity of the symptoms. In a study by Tampke and Irwin, the authors confirm these findings that dissociative processes and symptoms are predicted by anxiety [370].

There is a higher presence of dissociative symptoms forming the association rules in this analysis. This is interesting to explore these high occurrences due to a large amount of research that supports these relationships. For instance, Bremner et al. found there was a significantly higher level of dissociative symptoms in patients with PTSD than in patients without PTSD. Dissociative symptoms are an important element of the long-term psychopathological response to trauma [366]. Rules No. 58-60 identify the concepts of distress and helplessness in the correspondence to amnesia. Gershuny et al. [367] theorized that dissociative phenomena and subsequent trauma-related distress may relate to fears about death or loss of control above and beyond the occurrence of the traumatic event itself. From examining these results, avoidance and depression contributed directly to a high quantity of PTSD symptom association rules.

Disassociation-related concepts forming associations is found to be the strongest correspondence with PTSD and a traumatic event.

5.6 Discussion

The most commonly identified PTSD symptoms among corpus concepts make up the co-occurrences of formed clusters with the highest probabilities. Findings from the cluster analysis show that concepts group consistently between clusters in accordance with their hierarchical representations from DSM diagnostic criterion when supported by PTSDO. The concepts group less consistently with respect to DSM when supported by the UMLS. Adjustments to minimum threshold display interesting relationships with reduced associative strength but merit further investigation.

Accuracy of PTSD diagnosis is precipitated by the definitive presence of a trauma or stressor, however it is often not identified by a healthcare provider because it is: a) not brought up by the patient b) not inquired by provider in assessment, or c) not identified in progress notes unless the trauma entails a physical injury or disease. When an important predictor of PTSD, such as a specific trauma is strictly available in narrative text, this research proposes identification of meaningful DSM clusters with increased granularity useful for potential clinical monitoring. By utilizing a few symptom association rules within the clusters, the monitoring can alert attention to the needed screening of a patient for PTSD diagnosis. In future expansion of these techniques, there is prospective meaningful insights to be gained by performing clustering analysis of the association rules generated as opposed to its co-occurrence of concepts.

A disadvantage of the co-occurrence measure is that it favors concept pairs with high frequency. These frequent co-occurrences will be prevalent more often than infrequent pairs even if they are unrelated [215]. Based upon the cluster analysis performed in this research, the diagnostic construct of PTSD does appear to accurately describe some features of a universal trauma response. Three of the clusters were made up of only two DSM criterion categories compared to setting the cluster count to eleven, as recommended by heuristics determination, had only two of the clusters consisting of two DSM criterion. With the six clustering of UMLS supported extractions, a single clustering of concepts is made up of a greater amount of non-applicable PTSD concepts. The clusters formed with concept extraction with PTSDO are slightly more similar than those supported with the UMLS which were more distinct clusters. The EM

algorithm does offer some advantage in comparison to k-means clustering. Its computation can be completed in one iteration of the corpus and requires less computing power. This is not necessarily helpful on a small corpus but as data grows and real-time processing becomes more relevant, this clustering algorithm could outperform other techniques. There are specific limitations to performing cluster analysis and criticisms are that it is often considered merely descriptive, theoretical, and non-inferential. Regardless of any value present in the data analyzed, clusters will be produced regardless of the actual existence of any structure. The cluster analysis is not always generalizable because it is completely dependent upon the variables used as a basis for the similarity measure.

Applying association rules in medical diagnosis can be used for assisting physicians to aid patients. The general problem of the induction of reliable diagnostic rules is hard because theoretically no induction process by itself can guarantee the correctness of induced hypotheses [219]. Practically, diagnosis is not an easy process as it involves unreliable diagnosis tests and the presence of noise in training examples. This may result in hypotheses with unsatisfactory prediction accuracy which is too unreliable for critical medical applications [219, 368]. Serban has proposed a technique based on relational association rules and supervised learning methods. It helps to identify the probability of illness in a certain disease. This interface can be simply extended by adding new symptoms types for the given disease, and by defining new relations between these symptoms [219].

Findings with this process is rich in dissociation related concepts corresponding to trauma exposure and posttraumatic stress. This research is consistent with presented published literature supporting these findings. There are many occurrences of PTSD and dissociative disorder concepts because the prevalence of these symptoms do overlap. A trauma can propagate great emotional and mental disruption causing an individual to dissociate as a coping mechanism. Taken together with this work, these studies suggest a need for additional research evaluating the temporal relationship between dissociation and PTSD [215, 366-367]. Several rules portray correspondence between dissociative and arousal concepts. Feeny et al. explored similar associations between anger and dissociation and their relationship to symptoms of post-trauma pathology. In this study, anger expression was predictive of later PTSD severity, whereas dissociation was predictive of poorer later functioning [371].

To explore hypothesis generating prospects of association rules, the findings of dissociative type symptoms appearing to frequently correspond to posttraumatic stress and trauma exposures is further investigated. Using a variant of Swanson's open discovery [372] where unknown or underreported relationships are mined in a process known as literature-based discovery [373]. To search for knowledge, the following process functions used by Weeber et al. [374] are explored: if "A is related to B" and "B is related to C", then the hypothesis that "A causes C" is strongly suggested. Found in the analyses are many dissociative concepts that corresponded to stressor concepts shown in the following rule:

	LHS	RHS	support	confidence	lift
Rule 1:	{amnesia, pain, detachment}	=> {trauma exposure}	0.29	0.65	12.62

Amnesia, pain, and detachment are concepts related to or thought to relate to disassociation. As discussed, the literature is full of studies confirming this dissociative concept link to PTSD and stressors. Among the large quantity of rules generated, the presence of symptom concepts that fall under the avoidance criterion in the DSM are abundant. What is not found is any minimal or specific concepts of avoidance related concepts with significant links to stressors or posttraumatic stress. However, there are many avoidance-based concepts that correspond to several dissociative-based concepts as shown in the following rule:

	LHS	RHS	support	confidence	lift
Rule 2:	{agoraphobia, lack engagement, blunted affect}	=> {amnesia}	0.24	0.59	9.72

Agoraphobia, lacking engagement, and blunted affect correspond to the concept of amnesia. Using Swanson's A to C method of discovery, one could insinuate that avoidance-based concepts through linkage of dissociative-based concepts do correspond to stressors such as trauma exposure. At the current minimum support threshold of 0.24, this was not the case. However, after lowering the threshold to a support of 0.21, several avoidance-based concepts do in fact relate to stressor concepts such as in the rule:

	LHS	RHS	support	confidence	lift
Rule 3:	{agoraphobia, lack engagement}	=> {trauma exposure}	0.21	0.44	2.61

There are several limitations to be recognized when considering methods to implement association rules into research techniques. The quantity of association rules can grow rapidly into a collection of unmanageable set of rules. This is especially true for frequency requirements that are set to a low threshold. Also, many of the rules returned are redundant and unhelpful for analyses. Lastly, the majority of the algorithms do not always return the results in a reasonable time. In a setting where findings and associations are needed quickly to point researchers in specified directions, these techniques would not be useful. Apriori association technique has been proven to be effective in finding various trends in healthcare data. However, not all the rules generated are meaningful with this association rule approach. These techniques of data mining are just tools that provide methods of generating hypotheses. It does not verify the hypothesis; nor does it provide any information regarding the value of the generated hypotheses. Further analyses and research is needed to verify and explore these findings in a more scientific approach.

5.7 Conclusion

Presented in this analysis is the application of common data mining techniques to a readily available corpora of medical data. Based upon the cluster analysis supported with the PTSDO performed in this research, the diagnostic construct of PTSD does appear to accurately describe some features of a universal trauma response. In particular, this research found the DSM avoidance cluster and the dissociative concepts it contains as a potential meaningful predictor of PTSD. The greater recall from concept extractions supported with the UMLS appears to affect the clustering of DSM criterion as well as the inclusion of concepts that do not pertain directly to PTSD symptomatology. Cluster analysis is not always generalizable and often considered merely descriptive, theoretical, and non-inferential. However, this method does identify specific important clusters that could be useful for future hypothesis generation. While this work would be considered too preliminary to suggest enhancements to diagnostic criteria, it does address some of the current criticisms of the clustering of symptoms in the DSM. It also demonstrates the value of the PTDSO as compared to the UMLS for clustering uniformly with DSM criterion and acquiring highly accurate associations.

PTSD has a unique position as the only psychiatric diagnosis that depends on a factor outside the individual, namely, a traumatic stressor [9]. Individuals with PTSD have an increased risk of

having other psychiatric disorders such as depression, other anxiety disorders, or substance abuse disorders. They are also more likely to have greater functional impairment, a reduced quality of life, and poorer physical health. Association rules are obtained that show relationships between various PTSD symptom concepts. They provide a method of measuring concept frequencies from extractions which provide potential to support diagnosis of PTSD from minimal symptomatology. These relationships point to intriguing questions of whether the feelings and emotions are legitimate PTSD sign or symptoms or forms of coping mechanisms.

Data mining plays a crucial role in mining of healthcare data. Healthcare data can be collected from various hospitals. The assimilated data can be used to analyze the patient reports which help in identifying the patterns present in the databases. This further helps to get information about diseases, their symptoms, causes, remedies and precautions that can help to prevent the occurrence of various diseases. This study shows that based only on this data set of easily attainable PubMed case reports, applying data mining techniques supported with a thorough concept identification provides meaningful insight. However, these findings are made obtainable from building a domain application-based terminology framework created for this research, PTSDO. This process can be extended to similar mental health domains where the range of symptoms complicate understanding. The cluster and association analysis can point to s diagnosis or a clinical screening alert for a disorder using minimal phenotypical information. Further exploration of these types of techniques have the potential to benefit medical informatics and researchers. In the future, it could be possible to apply the data mining methods to a richer, more clinical-based data set. In conclusion, these findings contribute to the long-term understanding of the interrelationships among PTSD symptoms. Additional factor analytic studies are needed to improve our understanding of the PTSD symptom structure and the stability of that structure over time.

Chapter 6

Discussion of Findings

6.1 Synthesis of Findings

The growing scientific evidence of PTSD information is multiplying at a rapid rate which includes genetic biomarkers, risk factors, brain chemistry output, and phenotypical narrative text. Independent databases of this knowledge are expanding creating opportunities to engineer interoperable methods to acquire information from these resources with regards to terminological synthesis, text and data mining techniques. After thorough stakeholder analysis, requirements gathering, and knowledge acquisition it is determined that existing resources do not contain sufficient resources in order to acquire salient terminology. Traditional means of acquiring knowledge from existing biomedical terminologies and domain expert interviews do not provide the necessary augmentation for domain coverage expansion. Developing requirements is found to be the greatest challenges in terminology development. The lack of understanding of what is offered from comprehensive domain terminologies is highly prevalent. Substantial progress was consistently impeded until a project champion provided support for overcoming misunderstandings between stakeholders. The majority of the terminology system requirements identified addressed terminology system design, concept inclusion, and text-mining interoperability support.

There were several iterations of revising needs and requirements that slowed terminology development from the traditional gathering methods. The use of agile methodological processes is invaluable to the success of this research allowing the recognition of errors, fixing them, and continuing on to the next problem. The techniques provided by the method allows for error correction quickly while not impacting the momentum of the entire project. The majority of stakeholders that provided guidance to the research are DoD employees and with the Veterans Healthcare system which greatly influenced the purpose and scope of terminology development. Modeling symptoms according to DSM arrangement and treatments corresponding to VA/DoD guidelines was mandated despite stakeholder acknowledgement that this would inhibit

community user acceptance outside of federal healthcare. Domain modeling complications also arose due to the many comorbid mental health symptoms that were collected during knowledge acquisition.

Semi-automated techniques such as those described in this dissertation are necessary in order to supplement knowledge acquisition. This is shown in the high number of de novo concepts and terms created in order to adequately describe the domain. The definition development for these de novo concepts are complex as there is great misinterpretation among terms in the disorder and each requires much granularity for maximizing recall. The semi-automated techniques do reach a point of saturation for acquiring new knowledge, however these steps will need to be repeated as new knowledge is discovered or comes available in order to expand PTSDO. Many of the automated tools designed to aid in this process are not user friendly, provide incomplete information, or not interoperable with all relevant terminology resources. There are several difficulties in found in the knowledge acquisition process for terminology development. Locating existing resources is time-consuming and much of the knowledge obtained can be ambiguous, overlapping, and contradictory to knowledge already incorporated. Coordinating the availability of stakeholders in order to obtain expert feedback is constantly paramount. Communication management is required to gather relevant feedback as stakeholders may be knowledgeable but lack the ability to relay salient information. Verifying and validating the data once acquired can be complex as well. Lastly, developing automated knowledge gathering is important as the resources of salient information continues to multiply as the project progresses. New resources of new electronic data become available for mining as new stakeholders become involved and as interest grows. The automated processes aid in filtering of noisy and irrelevant data that permeates the growing information overload.

Extracting the entities within this disorder by clinicians, curators, researchers, and other stakeholders takes considerable time and resources. Text mining initiatives ease these processes and improves performance of developing these clinical applications. The development of gold standards is necessary for the training and evaluation of text mining pipelines, however the annotation process is very laborious and time-consuming. The inter-annotator agreement achieved in this dissertation is excellent although the annotators and adjudicator are domain experts or at least legitimate training. Without their expertise, it is proposed that reasonable annotation agreement could be achieved with significant attention to guideline development and

maximum training sessions. This gold standard creation was hampered when changing stakeholder requirements for the terminology development affected the annotation schema and tasks. While this corpora of annotated data are difficult to build, a benefit is that it is easily shareable to a multitude of text mining tasks and its methods and their respective evaluation can improve other biomedical applications.

The recognition of PTSD concepts significantly varies in accuracy among terminology resources for supporting several text mining pipelines when processing biomedical corpora. Many existing terminologies are missing PTSD-related concepts and terms which do not support dictionary-lookup features for identification. Additionally, the non-specificity in the terminologies fosters the identification of spurious concepts not applicable to the domain of the disorder. Error analysis reveals various reasons and types of error occurrence consisting of the following categories: a) incorrect sub-domain; b) pipeline deficiencies; c) annotation mistakes; d) textual misidentification; e) semantic type errors; and f) not available in dictionary.

The majority of false-negatives is due to concepts simply not existing in a dictionary, thus are not identified due to no available reference for the dictionary-lookup algorithm. Text mining deficiencies such as tokenization, stemming, normalization, and word sense disambiguation, etc. caused inaccuracies beyond the control of the terminology reference. Specific semantic types are removed in order to reduce the high number of false-positive errors, however, this limitation does produce false-negatives. Errors also derive from textual misidentification such as misspellings in the text, negation errors, and missed identification due to acronyms and abbreviations. Those produced from lack of dictionary references are correctable with the methods described within this dissertation. Pipeline deficiencies are also correctable however, every system will typically be prone to some NLP shortcomings depending upon the needed tasks. Spelling mistakes are found but these errors are unavoidable and inherent in most narrative or transcribed text.

Dictionary-based terminology resources do not perform equally on corpora with text processing pipelines. The accuracy of dictionary coverage varies significantly between many of the text processing pipelines. The task of keeping up with comprehensive knowledge coverage resources will continue to be a time-consuming limitation for maintaining accuracy. As new knowledge is inserted within the domain, it must be identified and included in the terminology. The more automation that can be implemented into these tasks, the more researchers and

developers can maintain efficiency. Findings from the cluster analysis show that concepts group uniformly with DSM diagnostic criterion when supported with PTSDO. The concepts group less uniformly with respect to DSM when supported by the UMLS. However, the cluster analysis is not always generalizable because it is completely dependent upon the variables used as a basis for the similarity measure. The association rules developed with the extracted concepts provide useful insights into correspondence between PTSD symptomatology. Analysis of association rule mining supports hypothesis generation via the identification of relationships between symptoms that merit further investigation.

Findings with the systems development and algorithm implementations discover a high-level of dissociation-related concepts corresponding to trauma exposure and posttraumatic stress. Among the large quantity of rules generated, the presence of symptom concepts that fall under the avoidance criterion in the DSM are abundant. What is not found is specific concepts of avoidance-related concepts with significant links to stressors or posttraumatic stress. However, there are many avoidance-based concepts that correspond to several dissociative-based concepts. Association rule mining in this research implemented a variant of Swanson's open discovery to explore hypothesis generating prospects of these dissociative type symptoms. After lowering the support threshold for exploration purposes and using Swanson's A to C method of discovery, it could be insinuated that avoidance-based concepts through linkage of dissociative-based concepts do in fact correspond to stressors such as trauma exposure. This example exhibits the usefulness of association rule mining for potential monitoring of individuals at risk for developing PTSD. Further analyses and research is needed to verify and explore these findings in a more scientific approach.

Keeping current with the latest clinical and scientific findings of PTSD and in the field of mental health is a formidable task. The growing rate of information requires automated techniques to advance in order to allow researchers and clinicians to keep pace. The development of domain-specific terminologies, implementation of text mining pipelines, and pattern identification with data mining algorithms provide novel opportunities to address information overload concerns logically and thoroughly.

6.2 Future Work

Future applications will continue to need terminology support for text mining initiatives, artificial intelligence applications, and qualitative analysis to implement accurate clinical decision support systems. Decision-making processes for PTSD are complex and heavily knowledge based, requiring analysis of multiple items of information. Expanding this work on more advanced terminology systems can foster greater community collaboration providing opportunities for overcoming knowledge gaps and improve knowledge discovery for this complex and prevalent clinical disorder. Additionally, these methods and the framework applied in this research is extendable to other mental health disorders and other medical conditions.

There are growing collections of electronic data such as web forums, literature, and EHR progress notes providing means to support victims and sufferers on a greater scale. It is important to examine the similarities and disparities that exist between biomedical text and clinical text. The concept recognition output of this dissertation can be applied to a greater number of use cases when analyzing additional textual documentation as it comes available. There are opportunities to apply the methods presented in this dissertation to address challenges of co-morbidity and entity extraction annotated with temporality features. The continued development of PTSDO must support greater database integration through use of linked data and mapped terminologies. Term reuse within the development community must be addressed with greater urgency as considerations can reduce engineering costs and support semantic interoperability among different datasets and applications. The great amount of overlapping terms with minimal reuse in comparison goes against the purpose of terminology standardization, interoperability, and sharing that needs to be pursued much more aggressively in development initiatives. There is also an opportunity to develop more robust annotation tools that are much more user friendly, specifically for identification of named entities.

PTSDO is built with top-level ontology characteristics to support future enhancements towards a PTSD ontology. Its development can support improved patient care by advancing research and reducing redundancy or misinformation. Developing a PTSD ontology will be significant because it will organize disparate data, enhance the clinical understanding of the disorder, and increase the knowledge of prediction, prognosis and recovery (e.g., longitudinal tracking or natural history of the disorder) of patients. As text mining initiatives advance and data science continues to move to the forefront of biomedical support, ontologies have great

potential to address data-driven research challenges. For acceptance, the ontology must receive input from, and use by, a diverse community of experts and stakeholders across many domains of biomedicine.

Chapter 7

Conclusion

7.1 Summary and Lessons Learned

With the growing prevalence of mental illness, novel initiatives such as the tasks in this dissertation are needed to explore means to assist in improved understanding of disorders. The subjective nature of clinical diagnoses and symptom identification is problematic for many disorders including PTSD. The comorbidity of symptoms is an obstacle to explicit identification of concepts for very similar disorders. Concentrating focus on improving the terminology and systems to describe PTSD and these other disorders provides an opportunity to contribute long-term support regardless of the future applications that researchers will develop.

The heterogeneity of data between various databases and applications must be considered for achieving absolute information interoperability. Several pipelines explored in this dissertation are created in the UIMA framework which is the same NLP framework that forms the backbone of Watson. Watson is IBM's artificial intelligence supercomputer capable of answering vast human questions and made famous with its success on the game show Jeopardy. It is capable of performing speech recognition, NLP, data mining, and reasoning with a high degree of accuracy. While medicine will likely not advance in the foreseeable future to make complicated clinical decisions without physician intervention, Watson highlights the possibilities of similar systems making useful suggestions to support clinicians. By supporting requirements that balance domain coverage with various stakeholders needs and system usability, PTSDO has promise to support future interface technologies with terminology support. The level of interoperability required varies with application needs bringing issues of usability and usability engineering to the forefront of future terminology development to support text and data mining.

The techniques applied in this research can not only discover information themselves but can also be vital for supporting other software development initiatives. Collections of annotated training data is and will increasingly be extremely important to the rapidly developing future needs of machine learning applications. Access to gold standards will continue to be a gap in

furthering the development of text mining and NLP until the annotation of multi-domain data is fully automated. The automated methods of annotation in this dissertation can directly support this development of training data. As applications and research advance, assertional knowledge must be reconciled against definitional knowledge related to PTSD. The information acquired will require validation by domain experts and by clinical studies. There is an opportunity to document explicit metadata within terminologies that support these applications which can aid in acquisition of new knowledge and clinical information advancement. The assertional facts are vital for categorizing information and implementing applications that can logically reason over data by automated technologies. The knowledge will directly impact terminology development.

The unstructured nature of the growing collection of biomedical data makes it difficult to extract knowledge from the plethora of heterogeneous sources. In order to prepare this collection for data mining and extraction initiatives, accurate annotation or markup of the text is a critical first requirement. Each strategically designed use case will determine the layer of annotation, whether lexical, syntactic, or semantically based. Many text mining systems and developers have not focused on the terminologies which they rely upon providing an opportunity to enhance processing within the research field. Systems that use efficient NLP with focus on NER have the greatest potential for building the most accurate data mining and advanced information retrieval systems of the future. Data mining techniques supported with advanced NLP coverage can enhance the processing power for clinical decision support systems by increasing knowledge action-based suggestions. Decision support can facilitate proactive preventive intervention delivering just-in-time, actionable knowledge. The future advancement of text mining systems will rely upon more semantically robust lexicons in order to address complex clinical questions. Greater processing power can be achieved with thorough semantically annotated corpora actively linked to source terminologies. Extracted data can provide useful information for mental health support at varying levels of natural language understanding tailoring it to individual patient needs. Terminologies must be evidenced-based and evolve gracefully as findings change and new information becomes available.

7.2 Closing Statement

Veterans are committing suicide at a rate of over 22 taken lives a day. The majority are confirmed sufferers of PTSD. Civilians deal with this disorder at staggering rates as well. Mental

health has not received the research focus that it deserves. This is due to unavoidable economic factors, attention needed to other serious health conditions, and lack of understanding the core impact brain health has on systems biology. As wars continue to manifest themselves around the globe, as natural disasters are unavoidable tragedies, and as humans continue to cause other humans harm, PTSD will continue to remain prevalent and impede patient's ability to focus on physical health conditions and address the root of their respective issues. Originally, the collection of symptoms was thought to develop from the impact of artillery shells and referred to as 'shell shock' but evolved to Freud's model of 'war neurosis' to include expanded etiological circumstances [372]. Research was greatly expanded with survivors of the Nazi Holocaust to include traumatic experiences from other man-made disasters, natural disasters, and assaults as stressors distinguishable from life disparities [8, 9]. Large sums of money are being spent, but with the increasing incidence and the high suicide-rate, research as currently conducted is questionable. Populations have been dealing with this disorder a long time, but unfortunately healthcare communities continue to make the same repeated mistakes. Communities have been dealing with this disorder a long time, and healthcare continues to make the same repeated mistakes. Future research must be more innovative and promote strategies that impact the disorder upstream for prediction and prevention. Continued initiatives with terminological support, text mining implementations, and data mining development will support overcoming the difficulty in describing the range of symptoms along with the wide array of possible treatments for the disorder. In conclusion, knowledge about the epidemiology of PTSD is important for researchers to help guide scientific inquiry, and for clinicians to help them gain greater understanding of their patients and use this understanding to enhance treatment outcome. Knowledge resides in the abundant electronic information overload and big data problem space. Novel approaches as those described in this dissertation will provide the proficiencies to acquire the knowledge and garner meaning.

REFERENCES

1. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public health reports 1985;100(2):126.

2. Bornman, J. The World Health Organisation's terminology and classification: application to severe disability. Disability and rehabilitation 2004;26(3):182-8.

3. Lundberg C et al. Selecting a standardized terminology for the electronic health record that reveals the impact of nursing on patient care. **Online** J Nurs Inform 2008;12(2).

4. Timmermans S, Berg M. The gold standard: The challenge of evidence-based medicine and standardization in health care. Philadelphia: Temple University Press; 2010.

5. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th ed.) Washington, DC; 1994.

6. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th ed.) Washington, DC; 2013.

7. Norris F, Tracy M, Galea S. Looking for resilience: Understanding the longitudinal trajectories of responses to stress. Soc Sci Med 2009;68(12):2190-8.

8. Friedman MJ, Keane TM, Resick PA (Eds.). Handbook of PTSD: Science and practice. New York: Guilford Press, 2007.

9. Gradus JL. PTSD: National Center for PTSD. Epidemiology of PTSD, Washington, DC: U.S. Department of Veterans Affairs 2014;(1).

10. Ramchand R, Schell TL, Karney BR, Osilla KC, Burns RM, Caldarone LB. Disparate prevalence estimates of PTSD among service members who served in Iraq and Afghanistan: possible explanations. J Trauma Stress 2010;23(1):59-68.

11. Teich JM, Osheroff JA, Pifer EA, Sittig DF, Jenders RA. Clinical decision support in electronic prescribing: recommendations and an action plan. J Am Med Inform Assoc 2005;12(4):365-76.

12. Robertson S, Robertson J. Mastering the requirements process: Getting requirements right. London: Addison-Wesley; 2012.

13. Sommerville I, Sawyer P. Requirements engineering: a good practice guide. New York: John Wiley & Sons, Inc.; 1997.

14. Hastings J, Smith B, Ceusters W, Jensen M, Mulligan K. Representing mental functioning: Ontologies for mental health and disease. In ICBO 2012: 3rd International Conference on Biomedical Ontology; 2012;7.

15. Freitas F, Schulz S, Moraes, E. Survey of current terminologies and ontologies in biology and medicine. RECIIS - Electronic Journal in Communication, Information and Innovation in Health 2009; 3(1):7–18.

16. Cohen AM, Hersh WR. A survey of current work in biomedical text mining. Brief Bioinform 2005;6(1):57-71.

17. Häyrinen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. Int J Med Inform 2008;77(5):291-304.

18. Manning CD, Schütze H. Foundations of statistical natural language processing. Cambridge: MIT press; 1999;(999).

19. Soderland S. Learning information extraction rules for semi-structured and free text. Mach Learn 1999;34(1-3):233-72.

20. Spackman KA, Campbell KE, Côté RA. SNOMED RT: a reference terminology for health care. Proc AMIA Symp 1997;640.

21. Elkin PL (Ed.). Terminology and terminological systems. Berlin: Springer Science & Business Media; 2012.

22. Nadkarni PM, Marenco L, Chen R, Skoufos E, Shepherd G, Miller P. Organization of heterogeneous scientific data using the EAV/CR representation. J Am Med Inform Assoc, 1999;6(6):478-93.

23. Rector AL. Clinical terminology: why is it so hard? Methods Inf Med 1999;38(4–5):239–52.

24. Nadkarni P, Chen R, Brandt C. UMLS concept indexing for production databases: a feasibility study. J Am Med Inform Assoc 2001;8(1):80–91.

25. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544-51.

26. Kardiner A. The traumatic neuroses of war. Washington, DC: National Academies; 1941.

27. Healy D. Images of Trauma: From Hysteria to Post-Traumatic Stress Disorder. London; Boston: Faber and Faber; 1993.

28. Wilson JP, Raphael B, eds. International handbook of traumatic stress syndromes. Berlin: Springer Science & Business Media; 2013.

29. Allen JG. Coping With Trauma: A Guide to Self-Understanding. American Psychiatric Press; 1995.

30. Peterson KC, Prout MF, Schwarz RA. Post-traumatic stress disorder: A clinician's guide. Berlin: Springer Science & Business Media; 2013;(6).

31. Freedy JR, Hobfoll SE, editors. Traumatic stress: From theory to practice. Berlin: Springer Science & Business Media; 2013 Jun 29.

32. Kleber RJ, Brom D, Defares PB. Coping with trauma: Theory, prevention and treatment. Netherlands: Swets & Zeitlinger Publishers; 1992.

33. Foy DW. Treating PTSD: Cognitive-Behavioral Strategies. New York: Guilford Press; 1992.

34. Department of Veterans Affairs. VA/DoD clinical practice guideline for post-traumatic stress management. Washington, DC: Department of Defense 2004; 2010 update.

35. Ursano RJ, McCaughey BG. Individual and community responses to trauma and disaster: The structure of human chaos. Cambridge: Cambridge University Press; 1995.

36. Beall LS. Post-traumatic stress disorder: A bibliographic essay. Choice 1997;34(6):917-30.

37. Helzer JE, Robins LN, McEvoy L. Post-traumatic stress disorder in the general population. N Engl J Med 1987;317(26):1630-4.

38. Yehuda R. Post-traumatic stress disorder. N Engl J Med 2002;346(2):108-14.

39. Dohrenwend BP, Turner JB, Turse NA, Adams BG, Koenen KC, Marshall R. The psychological risks of Vietnam for U.S. veterans: a revisit with new data and methods. Science 2006;313:979-82.

40. Goldberg J et al. Prevalence of Post-Traumatic Stress Disorder in Aging Vietnam-Era Veterans: Veterans Administration Cooperative Study 569: Course and Consequences of Post-Traumatic Stress Disorder in Vietnam-Era Veteran Twins. Am J Geriatr **Psychiatry** 20155(16).

41. Fulton JJ et al. The prevalence of posttraumatic stress disorder in operation enduring freedom/operation Iraqi freedom (OEF/OIF) veterans: a meta-analysis. J Anxiety Disord 2015;(31):98-107.

42. Warden DL, Ryan LM, Helmick KM, Schwab K, French L, Lu W, Ecklund J. War neurotrauma: The defense and veterans brain injury center (DVBIC) experience at Walter Reed Army Medical Center (WRAMC). Journal Neurotrauma 2005;22(10):1178.

43. Vanderploeg RD, Schwab K, Walker WC, Fraser JA, Sigford BJ, Date ES, Defense and Veterans Brain Injury Center Study Group. Rehabilitation of traumatic brain injury in active duty military personnel and veterans: Defense and Veterans Brain Injury Center randomized controlled trial of two rehabilitation approaches. Arch Phys Med Rehabil 2008;89(12):2227-38.

44. Menon DK, Schwab K, Wright DW, Maas AI. Position statement: definition of traumatic brain injury. Arch Phys Med Rehabil 2010;91(11):1637-40.

45. Jaffee CMS, Meyer KS. A brief overview of traumatic brain injury (TBI) and post-traumatic stress disorder (PTSD) within the Department of Defense. Clin Neuropsychol 2009;23(8):1291-8.

46. Gutierrez PM. Military Suicide Research Consortium (MSRC). Denver Research Institute Co; 2013;10:1.

47. Wilson JP, Keane TM. (Eds.). Assessing psychological trauma and PTSD. New York: Guilford Press; 2004.

48. Elhai JD, Gray MJ, Kashdan TB, Franklin CL. Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects?: A survey of traumatic stress professionals. J Trauma Stress 2005;18(5):541-45.

49. Foa EB, Keane TM, Friedman MJ, Cohen JA (Eds.). Effective treatments for PTSD: practice guidelines from the International Society for Traumatic Stress Studies. New York: Guilford Press; 2008.

50. Chapman WW, Nadkarni PM, Hirschman L, D'Avolio LW, Savova GK, Uzuner O. Overcoming barriers to NLP for clinical text: the role of shared tasks and the need for additional creative solutions. J Am Med Inform Assoc 2011;18(5):540-3.

51. Shiner B, D'Avolio LW, Nguyen TM, Zayed MH, Watts BV, Fiore L. Automated classification of psychotherapy note text: implications for quality assessment in PTSD care. J Eval Clin Pract 2012;18(3):698-701.

52. Stone K. NINDS common data element project: a long-awaited breakthrough in streamlining trials. Ann Neurol 2010;68(1):A11.

53. Thurmond VA., Hicks R, Gleason T, Miller AC, Szuflita N, Orman J, Schwab K. Advancing integrated research in psychological health and traumatic brain injury: common data elements. Arch Phys Med Rehabil 2010;91(11):1633-6.

54. Kaloupek DG, Chard KM, Freed MC, Peterson AL, Riggs DS, Stein MB, Tuma F. Common data elements for posttraumatic stress disorder research. Arch Phys Med Rehabil 2010; 91(11):1684-91.

55. Joyce C. Transforming Our Approach to Translational Neuroscience: The Role and Impact of Charitable Nonprofits in Research. Neuron 2014;84(3):526-32.

56. Wood FE, Fitzsimmons MJ. Clinical data interchange standards consortium (CDISC) standards and their implementation in a clinical data management system. Drug Inf J 2001;35(3):853-62.

57. Kaufman J. Healthcare and life sciences standards overview-Technology for Life: NC Symposium on Biotechnology and Bioinformatics. In Biotechnology and Bioinformatics, 2004. Proc Technology for Life: North Carolina Symp on IEEE 2004;10:31-41.

58. Spasic I, Ananiadou S, McNaught J, Kumar A. Text mining and ontologies in biomedicine: making sense of raw text. Brief Bioinform 2005;6(3):239-51.

59. Bernardi L, Ratsch E, Kania R. et al. Interdisciplinary work: The key to functional genomics, IEEE Intell Syst Trends Controv 2002;17(3):66–8.

60. Chang J, Schutze H, Altman R. Creating an online dictionary of abbreviations from Medline. J Am Med Inform Assoc 2002;9(6):612–20.

61. Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. Yearb Med Inform 2006:124-35.

62. Obrst L. Ontologies for semantically interoperable systems. In Proceedings of the twelfth international conference on Information and knowledge management. ACM 2003;(11):366-69.

63. World Health Organization. Manual of the International Classification of Diseases, Injuries, and causes of Death. Geneva, Switzerland; 1977.

64. Kilgarriff A, Yallop C. What's in a Thesaurus? In LREC 2000;(5).

65. Webster's ninth new collegiate dictionary. Springfield, MA: Merriam-Webster; 1990.

66. Jacquemin C. Spotting and discovering terms through natural language processing. Cambridge: MIT Press; 2001.

67. Kageura K, Umino B. Methods of automatic term recognition: A review. Terminology 1996;3(2):259-89.

68. Cimino JJ. From data to knowledge through concept-oriented terminologies. J Am Med Inform Assoc 2000;7(3):288-97.

69. Chute CG, Cohn SP, Campbell JR. A framework for comprehensive health terminology systems in the United States. J Am Med Inform Assoc 1998;(5):503–10.

70. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med 1998;37:394–403.

71. Berube MS, Neely DJ, DeVinne PB. The American heritage dictionary. Boston: Houghton Mifflin; 1982.

72. Armitage GC. Development of a classification system for periodontal diseases and conditions. Ann Periodontol 1999;4(1):1-6.

73. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985;13(10):818-29.

74. Gruber TR. A translation approach to portable ontologies. Knowl Acquis 1993;5(2):199-220.

75. Uschold M, Gruninger M. Ontologies: Principles, methods and applications. Knowl Eng Rev 1996;11(2):93-136.

76. Pinto SF, Martins JP. Ontologies: how can they be built? Knowl Inform Syst 2004;6: 441–64.

77. Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. Brief Bioinform 2006;7(3):256-74.

78. ISO 1087–1: Terminology work – vocabulary Part 1: theory and application: Technical Committee TC 37/SC 1; ISO Standards – terminology (principles and coordination); 1996.

79. ISO 1087–2: Terminology work – vocabulary Part 2: computer applications: Technical Committee TC 37/SC 3; ISO Standards – computer applications for terminology; 1996.

80. Rector AL, Nowlan WA. The GALEN project. Comput Methods Programs Biomed 1994;45(1-2):75-8.

81. Rogers J, Roberts A, Solomon D, Van der Haring E, Wroe C, Zanstra P, Rector A. GALEN ten years on: Tasks and supporting tools. Stud Health Technol Inform 2001;(1):256-60.

82. Rector AL, Rogers JE, Zanstra PE, Van Der Haring E. OpenGALEN: open source medical terminology and tools. Proc AMIA Symp 2003;982.

83. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The unified medical language system. J Am Med Inform Assoc 1998;5(1):1-11.

84. Rindflesch TC, Fiszman M, Libbus B. Semantic interpretation for the biomedical research literature. In Medical informatics, New York: Springer US; 2005; p. 399-422.

85. Saripalle RK. Current status of ontologies in Biomedical and Clinical informatics. University of Connecticut. 2004.

86. McCray AT, Srinivasan S, Browne AC. Lexical methods for managing variation in biomedical terminologies. Proc AMIA Symp 1994:235.

87. McCray AT. An upper-level ontology for the biomedical domain. CFG 2003;4(1):80-4.

88. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: The MetaMap program. Proc AMIA Symp 2001:17–21.

89. McCray AT, Bodenreider O, Malley JD, et al. Evaluating UMLS strings for natural language processing. Proc AMIA Symp 2001:448–52.

90. Lowe HJ, Barnett GO. Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. JAMA 1994;271(14):1103-8.

91. Lipscomb CE. Medical subject headings (MeSH). Bull Med Libr Assoc 2000;88(3):265.

92. Ashburner M et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000;25(1):25-9.

93. Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. Genome Res 2001;11(8):1425-33.

94. Elkin PL et al. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. In Mayo Clinic Proceedings. Elsevier, 2006;(81)1:741-8.

95. Botstein D et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000; 25(1):25-9.

96. Gene Ontology Consortium. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res 2004;(32)1:D258-61.

97. Donnelly K. SNOMED-CT: The advanced terminology and coding system for eHealth. Stud Health Technol Inform 2006;121:279.

98. Stearns MQ, Price C, Spackman KA, Wang AY. SNOMED clinical terms: overview of the development process and project status. Proc AMIA Symp 2001:662-6.

99. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. BMC Med Inform Decis Mak 2008;8(Suppl 1):S2.

100. de Coronado S, Haber MW, Sioutos N, Tuttle MS, Wright LW. NCI Thesaurus: using science-based terminology to integrate cancer research results. Medinfo 2004;11(Pt1):337.

101. Fragoso G, de Coronado S, Haber M, Hartel F, Wright L. Overview and utilization of the NCI thesaurus. CFG 2004;5(8):648-54.

102. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, Smith, B. The national center for biomedical ontology. J Am Med Inform Assoc 2012;19(2):190-5.

103. Musen MA. National Cancer Institute Thesaurus. Encyclopedia of Systems Biology; 2013;1492.

104. Forrey AW et al. Logical Observation Identifiers, Names and Codes (LOINC) database: a public use set of codes and names for electronic reporting of clinical laboratory test results. ClinChem 1996;42(1):81-90.

105. McDonald CJ et al. LOINC, a universal standard for identifying laboratory observations: a 5-year update. Clinical chemistry, 2003; 49(4): 624-633.

106. Huff SM et al. Development of the logical observation identifier names and codes (LOINC) vocabulary. J Am Med Inform Assoc 1998;5(3):276-92.

107. World Health Organization. International statistical classification of diseases and related health problems. World Health Organization; 2004.

108. World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.

109. Hasin D, Hatzenbuehler ML, Keyes K, Ogburn E. Substance use disorders: Diagnostic and Statistical Manual of Mental Disorders, (DSM-IV) and International Classification of Diseases, (ICD-10). Addiction 2006;101(s1):59-75.

110. Maercker A et al. Proposals for mental disorders specifically associated with stress in the International Classification of Diseases-11. Lancet 2013;381(9878):1683-5.

111. American Medical Association. CPT 1999: Current Procedural Terminology. New York: Aspen Publishers; 1995.

112. King MS, Sharp L, Lipsky MS. Accuracy of CPT evaluation and management coding by family physicians. Journal Am Board Fam Pract 2001;14(3):184-92.

113. Griffith HM, Robinson KR. Current Procedural Terminology (CPT) coded services provided by nurse specialists. Image Journal Nurs Sch 1993;25(3):178-86.

114. King MS, Lipsky MS, Sharp L. Expert agreement in Current Procedural Terminology evaluation and management coding. Arch Intern Med 2002;162(3):316-20.

115. Sands JK. Nursing Diagnoses: Definitions and Classification 2001–2002. Fam Community Health 2002;24(4):57-8.

116. Herdman TH (Ed.). Nursing Diagnoses 2012-14: Definitions and Classification. New York: John Wiley & Sons; 2011.

117. Bulechek GM, Butcher HK, Dochterman JMM, Wagner C. Nursing interventions classification (NIC). Philadelphia: Elsevier Health Sciences; 2013.

118. Moorhead S. Nursing outcomes classification (NOC). Encyclopedia for nursing research; 2006.

119. Henry SB, Mead CN. Nursing classification systems. J Am Med Inform Assoc 1997;4(3):222-32.

120. Warren JJ, Coenen A. International Classification for Nursing Practice (ICNP). J Am Med Inform Assoc 1998;5(4):335-6.

121. Bowker GC, Star SL. Sorting things out: Classification and its consequences. Cambridge: MIT Press; 2000.

122. Regier DA, Narrow WE, Kuhl EA, Kupfer DJ. The conceptual development of DSM-V. Am J Psychiatry; 2009; Jun 1.

123. Campbell RJ. Psychiatric dictionary. Oxford: Oxford University Press; 1989.

124. Banks JL. PILOTS (published international literature on traumatic stress) database. J Trauma Stress 1995;8(3):495-7.

125. Butler K. PILOTS (Published International Literature on Traumatic Stress). TCA 2015;16(3):23-4.

126. Bedard M, Greif JL, Buckley TC. International publication trends in the traumatic stress literature. J Trauma Stress 2004;17(2):97-101.

127. Hadzic M, Chen M, Dillon TS. Towards the mental health ontology. In Bioinformatics and Biomedicine. BIBM'08. IEEE International Conference on IEEE 2008;11: 284-8.

128. Smith B et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol 2007;25(11):1251–5.

129. Grenon P. Smith B. SNAP and SPAN: Towards dynamic spatial ontology. Spat Cogn Comput 2004;4(1):69–104.

130. Hastings J, Schulz S. Ontologies for human behavior analysis and their application to clinical data. Int Rev Neurobiol 2012;103: 89-107.

131. Ceusters W. Smith B. Foundations for a realist ontology of mental disease. J Biomed Semantics 2010a;1(1):10.

132. Khoozani EN, Hadzic M. Designing the human stress ontology: A formal framework to capture and represent knowledge about human stress. Australian Psychologist 2010;45(4):258-273.

133. Hastings J, Ceusters W, Smith B, Mulligan K. The emotion ontology: enabling interdisciplinary research in the affective sciences. In Modeling and Using Context. Berlin Heidelberg: Springer; 2011. p. 119-23.

134. Uschold M, King M. Towards a methodology for building ontologies. In Workshop on Basic Ontological Issues in Knowledge Sharing, held in conjunction with IJCAI-95, Montreal, Canada; 1995.

135. Bada M, Hunter LE, Eckert M, Palmer M. An overview of the CRAFT concept annotation guidelines. In Proceedings of the Fourth Linguistic Annotation Workshop: Association for Computational Linguistics; 2010:207–11.

136. Lu Z, Bada M, Ogren PV, Cohen KB, Hunter L. Improving biomedical corpus annotation guidelines. Proceedings of the Joint BioLINK and 9th Bio-Ontologies Meeting. Fortaleza, Brazil 2006;(1):89-92.

137. Rosse C, Mejino JL, Modayur BR, Jakobovits R, Hinshaw KP, Brinkley JF. Motivation and Organizational Principles for Anatomical Knowledge Representation The Digital Anatomist Symbolic Knowledge Base. J Am Med Inform Assoc 1998;5(1):17-40.

138. Noy N, McGuinness D. Ontology development 101: a guide to creating your first ontology, Stanford knowledge systems laboratory technical report KSL-01-05 and Stanford medical informatics technical report SMI-2001-0880; 2001.

139. Pérez AG, de Figueroa Baonza MCS, Villazón B. Neon methodology for building ontology networks: Ontology specification. Methodology 2008;1-18.

140. Suárez-Figueroa MC, Gómez-Pérez A, Villazón-Terrazas B. How to write and use the ontology requirements specification document. In On the move to meaningful internet systems: OTM 2009 Berlin Heidelberg: Springer; 2009. p. 966-82.

141. Grüninger, M, Fox M. Methodology for the Design and Evaluation of Ontologies. Proc. of IJCAI95's Workshop on Basic Ontological Issues in Knowledge Sharing; 1995.

142. Fernández M. et al. METHONTOLOGY: From Ontological Art Towards Ontological Engineering. Proc. AAAI Spring Symp. Series. AAAI Press, Menlo Park, Calif., 1997; 33-40.

143. Nenadić G, Mima H, Spasić I, Ananiadou S, Tsujii JI. Terminology-driven literature mining and knowledge acquisition in biomedicine. Int J Medical Inform 2002;67(1):33-48.

144. Hart A. Knowledge acquisition for expert systems. London: Springer; 1988. p. 103-11.

145. Hearst MA. Untangling text data mining. Proc 37th Annual meeting of the Association for Computational Linguistics. College Park, MD. 1999;(2):3-10.

146. Meystre SM, Savova GK, Kipper-Schuler KC, Hurdle JF. Extracting information from textual documents in the electronic health record: a review of recent research. Yearb Med Inform 2008;35:128-44.

147. Bird S, Klein E, Loper E. Natural language processing with Python. Sebastopol, CA: O'Reilly Media, Inc; 2009.

148. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994;1(2):161.

149. Leroy G, Chen H, Martinez JD. A shallow parser based on closed-class words to capture relations in biomedical text. J Biomed Inform 2003;36(3):145-58.

150. Grishman R, Sundheim B. Message understanding conference - 6: A brief history. In Proceedings of the 16th International Conference on Computational Linguistics (COLING); 1996.

151. McNaught J, Black WJ. Information extraction: the task. In: Ananiadou S, McNaught J, editors. Text Mining for Biology and Biomedicine: Artech House Books; 2006; 143-76.

152. Hobbs JR. The generic information extraction system. Proc MUC-5; Baltimore, MD: Morgan Kaufmann; 1993;87-92.

153. Hobbs JR. Information extraction from biomedical text. J Biomed Inform 2002;35(4):260-4.

154. Pakhomov S, Buntrock J, Duffy PH. High Throughput Modularized NLP System for Clinical Text 43rd Annual Meeting of the Association for Computational Linguistics; 2005; Ann Arbor, MI; 2005.

155. Hahn U, Romacker M, Schulz S. Creating knowledge repositories from biomedical reports: the MEDSYNDIKATE text mining system. Pac Symp Biocomput 2002:338-49.

156. Sager N, Friedman C, Chi E. The analysis and processing of clinical narrative. In: Salamon R, Blum B, Jørgensen M, editors. Medinfo 86; 1986; Amsterdam (Holland): Elsevier; 1986;1101-5.

157. Friedman C, Johnson SB, Forman B, Starren J. Architectural requirements for a multipurpose natural language processor in the clinical environment. Proc Annu Symp Comput Appl Med Care.1995:347-51.

158. Hripcsak G, Kuperman GJ, Friedman C. Extracting findings from narrative reports: software transferability and sources of physician disagreement. Methods Inf Med 1998;1-7.

159. Haug PJ, Ranum DL, Frederick PR. Computerized extraction of coded findings from free-text radiologic reports. Work in progress. Radiology 1990;174(2):543-8.

160. Haug PJ, Koehler S, Lau LM, Wang P, Rocha R, Huff SM. Experience with a mixed semantic/syntactic parser. Proc Annu Symp Comput Appl Med Care 1995;284-8.

161. McCray AT, Sponsler JL, Brylawski B, Browne AC. The role of lexical knowledge in biomedical text understanding. SCAMC 87; IEEE; 1987;103-7.

162. Lindberg C. The Unified Medical Language System (UMLS) of the National Library of Medicine. Journal (American Medical Record Association). 1990;61(5):40-2.

163. Ananiadou S, McNaught J. Text Mining for Biology and Biomedicine. Norwood, MA: Artech House, Inc; 2006.

164. Nenadic G, Spasic I, Ananiadou S. Mining biomedical abstracts: What is in a term?, in 'Natural Language Processing – IJCNLP 2004', Su KY et al., Eds, LNAI 3248, Berlin: Springer; 2004; 797–806.

165. Krauthammer M, Nenadic G. Term identification in the biomedical literature. J Biomed Inform 2004;37(6):512–26.

166. Kim J.-D, Otha T, Yoshimasa T, Yuka T, Collier N. Introduction to the bio-entity recognition task at JNLPBA. In Proceedings of the Joint Workshop on Natural Language Processing in Biomedicine and its Applications(JNLPBA-2004), Geneva, Switzerland; 2004.

167. Liu H, Wu ST, Li D, Jonnalagadda S, Sohn S, Wagholikar K, Chute CG. Towards a semantic lexicon for clinical natural language processing. Proc AMIA Symp 2012:568.

168. Chiang JH, Yu HC. Meke: Discovering the functions of gene products from biomedical literature via sentence alignment, BMC Bioinformatics 2003;19(11):1417–22.

169. Ferrucci D, Lally A. UIMA: an architectural approach to unstructured information processing in the corporate research environment. Nat Lang Eng 2004;10(3-4):327-48.

170. Manning CD, Surdeanu M, Bauer J, Finkel JR, Bethard S, McClosky D. The Stanford CoreNLP Natural Language Processing Toolkit. In ACL (System Demonstrations) 2014;6(1):55-60.

171. Cunningham H. GATE, a general architecture for text engineering. Computers and the Humanities 2002;36(2):223-54.

172. Maynard D, Tablan V, Cunningham H, Ursu C, Saggion H, Bontcheva K, Wilks Y. Architectural elements of language engineering robustness. Nat Lang Eng 2002;8(2-3):257-74.

173. Bird S. NLTK: the natural language toolkit. In Proceedings of the COLING/ACL on Interactive presentation sessions. Association for Computational Linguistics 2006; 7:69-72.

174. Baldridge J. The opennlp project. URL: http://opennlp. apache. org/index.html, accessed September 21, 2015.

175. Widdows D, Cohen T. The semantic vectors package: New algorithms and public tools for distributional semantics. In Semantic Computing (ICSC), 2010 IEEE Fourth International Conference on IEEE 2010;9:9-15.

176. McCandless M, Hatcher E, Gospodnetic O. Lucene in Action: Covers Apache Lucene 3.0. Greeenwich, CT: Manning Publications Co; 2010.

177. McCallum AK. Mallet: A machine learning for language toolkit; 2002

178. Bond F, Oepen S, Siegel M, Copestake A, Flickinger D. Open source machine translation with DELPH-IN. In Proceedings of the Open-Source Machine Translation Workshop at the 10th Machine Translation Summit 2005;9: 15-22.

179. Bond F, Nichols E, Appling DS, Paul M. Improving statistical machine translation by paraphrasing the training data. In IWSLT 2008;1: 150-7.

180. Verspoor K, Baumgartner Jr WA. Unstructured Information Management Architecture (UIMA). In Encyclopedia of Systems Biology. New York: Springer; 2013. p. 2320-24.

181. Hirschman L, Sager N, Lyman M. Automatic application of health care criteria to narrative patient records. Proc AMIA Symp on Computer Application in Medical Care 1979;10 :105

182. Sager N, Lyman M, et al. Natural language processing and the representation of clinical data. J Am Med Inform Assoc 1994;1(2):142–60.

183. Friedman C, Anderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language processor for clinical radiology. J Am Med Inform Assoc 1994;1:161–74.

184. Friedman C, Cimino JJ, Johnson SB. A schema for representing medical language applied to clinical radiology. J Am Med Inform Assoc 1994;1:233–48

185. Chiang JH, Lin JW, Yang CW. Automated evaluation of electronic discharge notes to assess quality of care for cardiovascular diseases using Medical Language Extraction and Encoding System (MedLEE). J Am Med Inform Assoc 2010;17(3):245-52.

186. Jain NL, Knirsch CA, Friedman C, Hripcsak G. Identification of suspected tuberculosis patients based on natural language processing of chest radiograph reports. Proc AMIA Symp 1996:542.

187. Zeng QT, Goryachev S, Weiss S, Sordo M, Murphy SN, Lazarus R. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. BMC Med Inform Decis Mak 2006; (6):30.

188. Jungermann F. Information extraction with rapidminer. In Proceedings of the GSCL Symposium' Sprachtechnologie und eHumanities 2009; 2:50-61.

189. Hofmann M, Klinkenberg R (Eds.). RapidMiner: Data mining use cases and business analytics applications. Boca Raton, FL: CRC Press; 2013.

190. Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-36.

191. Divita G, Tse T, Roth L. Failure analysis of MetaMap transfer (MMTx). Medinfo 2004;11(2):763-7.

192. Collier N, Oellrich A, Groza T. Concept selection for phenotypes and disease-related annotations using support vector machines. Proc. PhenoDay and Bio-Ontologies at ISMB; 2014.

193. Jonquet C, Shah NH, Musen MA: The open biomedical annotator. Summit Transl Bioinformatics 2009; 1:56.

194. Musen MA, Noy NF, Shah NH, Whetzel PL, Chute CG, Story MA, Smith B. The national center for biomedical ontology. J Am Med Inform Assoc 2012;19(2):190-5.

195. Whetzel PL, Noy NF, Shah NH, Alexander PR, Nyulas C, Tudorache T, Musen MA. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res 2011;39(suppl 2):W541-5.

196. Rindflesch TC, Fiszman M. The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform 2003;36(6):462-77.

197. Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth Inform Med 1993;32: 281-91.

198. Rosemblat G, Shin D, Kilicoglu H, Sneiderman C, Rindflesch TC. A methodology for extending domain coverage in SemRep. J Biomed Inform 2013;46(6):1099-107.

199. Hristovski D, Friedman C, Rindflesch TC, Peterlin B. 2006. Exploiting semantic relations for literature-based discovery. Proc AMIA Symp 2006:349-53.

200. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, Chute CG. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-13.

201. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A simple algorithm for identifying negated findings and diseases in discharge summaries. J Biomed Inform 2001:301-10.

202. Doan S, Conway M, Phuong TM, Ohno-Machado L. Natural language processing in biomedicine: a unified system architecture overview. Clinical Bio 2014; 8(1):275-294.

203. Resnik P. Semantic Similarity in a Taxonomy: An Information-Based Measure and its Application to Problems of Ambiguity in Natural Language. J Artif Intell Res 1999;11(1):95-130.

204. Hadzic M, Hadzic F, Dillon T. (, January). Tree mining in mental health domain. Proceedings of the 41st Annual International Conference on IEEE System Sciences; 2008 230.

205. Fodeh SJ, Zirkle M, Finch D, Reeves R, Erdos J, Brandt C. MedCat: A Framework for High Level Conceptualization of Medical Notes. Proceedings of the 13th International Conference on IEEE Data Mining Workshops 2013;12(1):274-80.

206. Witten IH, Frank E. Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann; 2005.

207. Epstein RJ. Unblocking blockbusters: using boolean text-mining to optimise clinical trial design and timeline for novel anticancer drugs. Cancer Inform 2009; 7:231–238.

208. Panagiotakopoulos T, Lyras D, Livaditis M, Sgarbas K, Anastasopoulos G, Lymberopoulos D. A contextual data mining approach towards assisting the treatment of anxiety disorders. IEEE Trans Inf Technol Biomed 2010;14(3):567-81.

209. Fayyad UM, Piatetsky-Shapiro G, Smyth P, Uthurusamy, R. Advances in knowledge discovery and data mining; 1996.

210. Tryon RC. Cumulative communality cluster analysis. Educational and psychological measurement; 1958.

211. Jain AK, Murty MN, Flynn PJ. Data clustering: a review. ACM computing surveys (CSUR) 1999;31(3):264-323.

212. Hartigan JA. Clustering algorithms. New York: John Wiley & Sons, Inc.; 1975.

213. Mirkin BG, Kramarenko AV. Approximate bicluster and tricluster boxes in the analysis of binary data. In Rough Sets, Fuzzy Sets, Data Mining and Granular Computing Berlin Heidelberg: Springer; 2011. p. 248-56.

214. Berkhin P. A survey of clustering data mining techniques. In Grouping multidimensional data. Berlin Heidelberg: Springer; 2006. p. 25-71.

215. Jain AK. Data clustering: 50 years beyond K-means. Pattern Recogn Lett 2010; 31(8), 651-666.

216. Momtazi S, Khudanpur S, Klakow D. A comparative study of word co-occurrence for term clustering in language model-based sentence retrieval. In Human Language Technologies: 2010 Annual Conference. Association for Computational Linguistics, 2010;(6): 325-328.

217. Brown PF, Pietra VJD, Souza PV, Lai JC, Mercer RL. Class-based n-gram models of natural language. Comput Linguist 1992;18(4):467–79.

218. Li H. Word clustering and disambiguation based on co-occurrence data. Nat Lang Eng 2002;8(01):25-42.

219. Zhang H, Korayem M, You E, Crandall DJ. Beyond co-occurrence: discovering and visualizing tag relationships from geo-spatial and temporal similarities. In Proceedings of the fifth ACM international conference on Web search and data mining. ACM 2012;(2): 33-42.

220. Cardie C. A case-based approach to knowledge acquisition for domain-specific sentence analysis. In 11th National Conference on Artifical Intelligence. Menlo Park, California: AAAI 1993;(1):798–803.

221. Ng HT. Exemplar-based word sense disambiguation: Some recent improvements. In C. Cardie & R. Weischedel (Eds.), Second Conference on Empirical Methods in Natural Language Processing (EMNLP-2). Somerset, New Jersey: Association for Computational Linguistics 1997;(1):208–13.

222. Zavrel J, Daelemans W. Memory-based learning: Using similarity for smoothing. In 35th Annual Meeting of the ACL. Somerset, New Jersey: Association for Computational Linguistics 1997;(1):436–43.

223. Rajak A, Gupta MK. Association rule mining-applications in various areas. In Proceedings of International Conference on Data Management, Ghaziabad, India. 2008; (2):3-7.

224. Jain D, Gautam S. Implementation of Apriori Algorithm in Health Care Sector: A Survey. Int J Computer Sci Eng Commum 2013;2(4).

225. Nuwangi SM, Oruthotaarachchi CR, Tilakaratna JMPP, Caldera HA. Utilization of Data Mining Techniques in Knowledge Extraction for Diminution of Diabetes, Second Vaagdevi International Conference on Information Technology for Real World Problems (VCON) 2010;(12):9-11.

226. Garner SR. WEKA: The Waikato Environment for Knowledge Analysis. In Proceedings of the New Zealand Computer Science Research Students Conference, 1995; 1:57–64.

227. Lekha A, Srikrishna CV, Viji V. Utility of Association Rule Mining: a Case Study using Weka Tool, 2013 International Conference on Emerging Trends in VLSI, Embedded System, Nano Electronics and Telecommunication System (ICEVENT) 2013;(1):7-9.

228. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD explorations newsletter 2009;11(1):10-18.

229. Bouckaert RR, Frank E, Hall MA, Holmes G, Pfahringer B, Reutemann P, Witten IH. WEKA-Experiences with a Java Open-Source Project. Journal Mach Learn Res 2010;11: 2533-41.

230. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Vanderplas J. Scikit-learn: Machine learning in Python. Journal Mach Learn Res 2011;12: 2825-30.

231. Van Der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. Comput Sci Eng 2011;13(2):22-30.

232. Jones E, Oliphant T, Peterson P. SciPy: Open source scientific tools for Python, 2001; http://www.scipy.org, accessed September 24, 2015.

233. Chang CC, Lin CJ. LIBSVM: a library for support vector machines. ACM Trans Intell Syst Technol (TIST) 2011;2(3):27.

234. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. Journal Mach Learn Res 2008;9: 1871–4.

235. Ganas S. Data Mining and Predictive Modeling with Excel 2007. In Casualty Actuarial Society E-Forum, Spring 2010.

236. Paul S, MacLennan J, Tang Z, Oveson S. Data Mining Tutorial. Microsoft Corporation, 2005; 50-59.

237. MacLennan J, Tang Z, Crivat B. Data mining with Microsoft SQL server 2008. New York: John Wiley & Sons; 2011.

238. Friedman MJ, Resick PA, Bryant RA, Brewin CR. Considering PTSD for DSM-5. Depress Anxiety 2011;28(9):750-69.

239. Ruggiero KJ, Rheingold AA, Resnick HS, Kilpatrick DG, Galea S. Comparison of two widely used PTSD-screening instruments: Implications for public mental health planning. J Trauma Stress, 2006;19(5):699-707.

240. Brown TA, Di Nardo PA, Lehman CL, Campbell LA. Reliability of DSM-IV anxiety and mood disorders: implications for the classification of emotional disorders. J Abnorm Psychol 2001;110(1):49.

241. Brewin CR, Lanius RA, Novac A, Schnyder U, Galea S. Reformulating PTSD for DSM-V: life after criterion A. J Trauma Stress 2009;22(5):366-73.

242. North CS, Suris AM, Davis M, Smith RP. Toward validation of the diagnosis of posttraumatic stress disorder. Am J Psychiatry 2009; 166:34–41

243. Karam EG et al. The role of criterion A2 in the DSM-IV diagnosis of posttraumatic stress disorder. Biol Psychiatry 2010;68(5):465–73.

244. McNally RJ. Progress and controversy in the study of posttraumatic stress disorder. Annu Rev Psychol 2003;54: 229–52.

245. Spitzer RL, First MB, Wakefield JC. Saving PTSD from itself in DSM-V. J Anxiety Disord 2007 21: 233–41.

246. Kirkbride JF. PTSD: An Elusive Definition. Journal of special operations medicine: a peer reviewed journal for SOF medical professionals 2011;12(2):42-7.

247. D'Avolio LW, Nguyen TM, Goryachev S, Fiore LD. Automated concept-level information extraction to reduce the need for custom software and rules development. J Am Med Inform Assoc 2011;18(5):607-13.

248. Zhu F et al. Biomedical text mining and its applications in cancer research. J Biomed Inform 2013;46(2):200-11.

249. Larsen P, von Ins M. The rate of growth in scientific publication and the decline in coverage provided by science citation index. Scientometrics 2010; 84:575-603.

250. Altman RB et al. Text mining for biology-the way forward: opinions from leading scientists. Genome Biol 2008;9(Suppl 2):S7.

251. Grimmer J, Stewart BM. Text as data: The promise and pitfalls of automatic content analysis methods for political texts. Polit Anal 2013;21(3):267-97.

252. Compton P, Jansen R. A philosophical basis for knowledge acquisition. Knowl Acquis 1990;2(3):241-58.

253. Puerta AR, Egar JW, Tu SW, Musen MA. A multiple-method knowledge-acquisition shell for the automatic generation of knowledge-acquisition tools. Knowl Acquis 1992;4(2):171-96.

254. Evans K, Herman SW. The National Center for PTSD. J Consum Health Internet 2014;18(1):81-8.

255. Kaloupek DG, Chard KM, Freed MC, Peterson AL, Riggs DS, Stein MB, Tuma F. Common data elements for posttraumatic stress disorder research. Arch Phys Med Rehabil 2010;91(11):1684-91.

256. Hunt RE, Newman RG. Medical knowledge overload: a disturbing trend for physicians. Health care Manage Rev 1997;22(1):70-5.

257. Orgun B, Vu J. HL7 ontology and mobile agents for interoperability in heterogeneous medical information systems. Comput Biology Med 2006;36(7):817-36.

258. Fung Y. Biomechanics: mechanical properties of living tissues. Berlin: Springer Science & Business Media; 2013.

259. Luther S, Berndt D, Finch D, Richardson M, Hickling E. Using statistical text mining to supplement the development of an ontology. J Biomed Info 2011;44: S86–93.

260. Kaufman RA, Rojas AM, Mayer H. Needs assessment: A user's guide. Educational Technology; 1993.

261. Gupta K. A practical guide to needs assessment. New York: John Wiley & Sons; 2011.

262. Wiesner S, Chen D, Ducq Y, Zanetti C. Methodology for Requirements Engineering & Evaluation in Manufacturing Service Ecosystems. Dissemination: Public Contributing to: WP 51;2012.

263. Nuseibeh B, Easterbrook S. Requirements engineering: a roadmap. In Proceedings of the Conference on the Future of Software Engineering. ACM 2000;5: 35-46.

264. Sören A, Auer S Herre H. RapidOWL—An Agile Knowledge Engineering Methodology. Perspectives of Systems Informatics. Berlin Heidelberg: Springer; 2006; (6):424-30.

265. Moore JW. Software engineering standards. New York: John Wiley & Sons, Inc.; 1998.

266. Abrahamsson P, Salo O, Ronkainen J, Warsta J. Agile software development methods: Review and analysis. EPSOO 2002, Finland: VTT Publications 2002;(2): 478.

267. Cohen D, Lindvall M, Costa P. Agile software development. DACS SOAR Report; 2003:11.

268. Beck K. Extreme programming explained: Embrace change. UK: Addison-Wesley; 2000.

269. Schwaber K, Beedle M. Agile software development with scrum. Australia: Prentice Hall PTR; 2001.

270. Stapleton J. DSDM −Dynamic system development method. UK: Addison-Wesley; 1995.

271. Highsmith JA. Adaptive software development. UK: Dorset House Publishing; 1996.

272. Cockburn A. Agile software development. London: Addison-Wesley; 2002.

273. Sillitti A, Giancarlo S. Requirements engineering for agile methods. Engineering and Managing Software Requirements. Berlin Heidelberg: Springer; 2005. p. 309-326.

274. Poppendieck T, Poppendieck M. Lean software development: An agile toolkit for software development managers. London: Addison-Wesley; 2003.

275. Glass R. Agile versus traditional: Make love, not war. Cutter IT Journal 2001;6(1):12−8.

276. Barrett JR, Jones DD. Knowledge engineering in agriculture. Amer Soc Agr Eng Monogr 1989.

277. Sowa JF. Knowledge representation: logical, philosophical, and computational foundations. Pacific Grove, CA: Brooks/Cole; 1999.

278. Sumaray A, Makki SK. A comparison of data serialization formats for optimal efficiency on a mobile platform. In Proceedings of the 6th international conference on ubiquitous information management and communication, ACM 2012; (2):48.

279. Antoniou G, Van Harmelen F. Web ontology language: Owl. In Handbook on ontologies Berlin Heidelberg: Springer. 2004. p. 67-92.

280. Smith MK, Welty C, McGuinness DL. OWL Web Ontology Language Guide, W3C Recommendation, 2004; http://www.w3.org/TR/2004/REC-owl-guide-20040210, accessed February 10, 2014.

281. Noy NF, Sintek M, Decker S, Crubezy M, Fergerson RW, Musen MA. Creating semantic web contents with protege-2000. IEEE Intelligent Systems 2001;16(2):60–71.

282. Knublauch H, Fergerson R, Noy N, Musen M. The Protégé´ OWL plugin: An open development environment for semantic web applications. In Third international semantic web conference 2004; 229–43.

283. Kola JS, Harris J, Lawrie S, Rector A, Goble C, Martone M. Towards an ontology for psychosis. Cogn Syst Res 2010;11(1):42-52.

284. Friedman C, Rindflesch TC, Corn M. Natural language processing: state of the art and prospects for significant progress, a workshop sponsored by the National Library of Medicine. J Biomed Inform 2013;46(5):765-73.

285. Fiszman M, Demner-Fushman D, Kilicoglu H, Rindflesch TC. Automatic summarization of MEDLINE citations for evidence-based medical treatment: A topic-oriented evaluation. J Biomed Inform 2009;42(5):801–13

286. Bodenreider O. The unified medical language system (UMLS): integrating biomedical terminology. Nucleic Acids Res 2004;32(suppl 1):D267-270.

287. Cabré Castellví MT. Theories of terminology: Their description, prescription and explanation. Terminology 2003;9(2):163-99.

288. Ivanović M, Budimac Z. An overview of ontologies and data resources in medical domains. Expert Syst Appl 2014;41(11):5158-66.

289. Duclos C, Burgun A, Lamy JB, Landais P, Rodrigues JM, Soualmia L, Zweigenbaum P. Medical Vocabulary, Terminological Resources and Information Coding in the Health Domain. In Medical Informatics, e-Health. Paris: Springer; 2014. p. 11-41.

290. ISO 704:2009. Terminology work - Principles and methods. International Organization for Standardization; 2009.

291. Kamdar MR, Tudorache T, Musen MA. A Systematic Analysis of Term Reuse and Term Overlap across Biomedical Ontologies. Semantic Web – Interoperability, Usability, Applicability (accepted - in press); 2009.

292. Witkin BR. Planning and conducting needs assessments: A practical guide. Thousand Oaks, CA: Sage; 1995.

293. Kizlik B. Measurement, assessment, and evaluation in education. ADPRIMA; 2012.

294. Sharp H, Finkelstein A, Galal G. Stakeholder identification in the requirements engineering process. In Database and Expert Systems Applications, 1999. Proceedings. Tenth International Workshop. IEEE 1999;(1): 387-91.

295. Boyce C, Neale P. Conducting in-depth interviews: A guide for designing and conducting in-depth interviews for evaluation input. Watertown, MA: Pathfinder International; 2006. p. 3-7.

296. De Vries H, Verheul H, Willemse H. Stakeholder identification in IT standardization processes. In Proceedings of the Workshop on Standard Making: A Critical Research Frontier for Information Systems. Seattle, WA; 2003;(12): 12-14.

297. Donaldson T, Preston LE. The stakeholder theory of the corporation: Concepts, evidence, and implications. Acad Manageme Rev 1995;20(1):65-91.

298. Freeman RE. Strategic Management: A stakeholder approach. Boston: Pitman; 1984.

299. Sharp H, Finkelstein A, Galal G. Stakeholder identification in the requirements engineering process. In Database and Expert Systems Applications, 1999. Proceedings. Tenth International Workshop on IEEE 1999;387-91.

300. Mitchell RK, Agle BR, Wood DJ. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Acad Manageme Rev 1997;22(4):853-86.

301. Young RR. Effective Requirements Practices. Boston, MA: Addison-Wesley; 2001.

302. Cheng BHC, Atlee JM. Research Directions in Requirements Engineering, Future of Software Engineering; 2007. p. 285-303.

303. Lopez MF, Gomez-Perez A, Sierra JP, Sierra AP. Building a chemical ontology using Methontology and the Ontology Design Environment. IEEE Intell Syst 1999;14(1):37–45.

304. Zhou L, Queen EB, Dongsong Z. ROD-toward rapid ontology development for underdeveloped domains. System Sciences, 2002. HICSS. Proceedings of the 35th Annual Hawaii International Conference on IEEE; 2002.

305. Doran GT. There's a S.M.A.R.T. way to write management's goals and objectives. Manage Rev 1981;70(11):35–6.

306. Bogue R. Use S.M.A.R.T. goals to launch management by objectives plan. TechRepublic. 2005; http://www.techrepublic.com/article/use-smart-goals-to-launch-management-by-objectives-plan/, accessed August 14, 2013.

307. Grüninger M, Fox MS. Methodology for the Design and Evaluation of Ontologies; 1995.

308. Cowie J, Lehnert W. Information extraction. Communications of the ACM 1996;39(1):80-91.

309. Cohen KB, Ogren PV, Fox L, Hunter L. Empirical Data on Corpus Design and Usage in Biomedical Natural Language Processing. Proc AMIA Symp 2005;156–160.

310. Cohen KB, Hunter L. Getting started in text mining. PLoS Comput Biol 2008;4(1):e20.

311. Eppler MJ, Mengis J. The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. Info Soc 2004;20(5):325-44.

312. Hall A, Walton G. Information overload within the health care system: a literature review. Health Info Librar J 2004;21(2):102-8.

313. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. J Info Science 2009;35(2):180-91.

314. Tan AH. Text mining: The state of the art and the challenges. In Proceedings of the PAKDD 1999 Workshop on Knowledge Discovery from Advanced Databases 1999;(4)8: 65-70.

315. Hearst MA. Text data mining: Issues, techniques, and the relationship to information access. Presentation notes for UW/MS workshop on data mining; July 1997.

316. Torrance MC. Natural communication with robots (Doctoral dissertation, Massachusetts Institute of Technology); 1957.

317. Cohn DA, Ghahramani Z, Jordan MI. Active learning with statistical models. J Artif Intell Res 1996.

318. Lehnert W, Cowie J. Evaluating an information extraction system, Commun. ACM 1994;39(1):80-91.

319. Kasabov N. Evolving Connectionist Systems: The Knowledge Engineering Approach, New York: Springer-Verlag, Inc.; 2006.

320. Hobbs JR, Riloff E. Information extraction. Handbook of natural language processing, Boca Raton, FL: CRC Press; 2010.

321. Nadeau D, Sekine S. A survey of named entity recognition and classification. Lingvist Invest 2007;30(1):3-26.

322. Maedche A, Staab S. Mining ontologies from text. In Knowledge Engineering and Knowledge Management Methods, Models, and Tools. Berlin Heidelberg: Springer; 2000. p. 189-202.

323. de Carvalho MG, Campos LM, Braganholo V, Campos MLA. Extracting new relations to improve ontology reuse. JIDM 2011;2(3):541.

324. Ananiadou S, Friedman C, Tsujii JI. Introduction: named entity recognition in biomedicine. J Biomed Inform 2004;37(6):393-5.

325. University of Sheffield Natural Language Group. Information Extraction: the GATE pipeline. 2011. http://www.gate.ac.uk/ie, accessed Jun 1, 2014.

326. Apache Foundation. The Unstructured Information Management Architecture. 2011. http://uima.apache.org, accessed Jun 3, 2014.

327. Browne AC, McCray AT, Srinivasan S. The SPECIALIST Lexicon 2000. http://lexsrv3.nlm.nih.gov/LexSysGroup/Projects/lexicon/2003/release/LEXICON/DOCS/techrpt.pdf, accessed Jul 20, 2014.

328. Divita G, Browne AC, Rindflesch TC. Evaluating lexical variant generation to improve information retrieval. Proc AMIA Symp 1998:775-9.

329. Koeling R, Carroll J, Tate AR, Nicholson A. Annotating a corpus of clinical text records for learning to recognize symptoms automatically. In Proceedings of the 3rd Louhi Workshop on Text and Data Mining of Health Documents, Bled, Slovenia. 2011;43-50.

330. Thompson P, Iqbal SA, McNaught J, Ananiadou S. Construction of an annotated corpus to support biomedical information extraction. BMC Bioinformatics 2009; 10:349.

331. Boisen S, Crystal MR, Schwartz R, Stone R, Weischedel R. Annotating resources for information extraction. In: Proceedings of the second language resources and evaluation. LREC 2000;1211–14.

332. Rost TB, Huseth O, Nytro O, Grimsmo A. Lessons from developing an annotated corpus of patient histories. J Comput Sci Eng 2008;2(2):162–79.

333. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc 2004;11(5):392–402.

334. Hunter L, LuZ, Firby J, Baumgartner WA, Johnson HL, Ogren PV, Cohen KB: Open DMAP: an open source, ontology-driven concept analysis engine, with applications to capturing knowledge regarding protein transport, protein interactions and cell-type-specific gene expression. BMC Bioinformatics 2008; 9:78.

335. Funk C, et al. Large-scale biomedical concept recognition: an evaluation of current automatic annotators and their parameters. BMC Bioinformatics 2014;15(1):59.

336. Smith MY, Redd WH, Peyser C, Vogl D. Post-traumatic stress disorder in cancer: a review. Psycho Oncology 1999;8(6):521-37.

337. Suomela BP, Andrade MA. Ranking the whole MEDLINE database according to a large training set using text indexing. BMC bioinformatics 2005;(6)1:1.

338. ESS, STREET PR. Counseling and psychotherapy transcripts, client narratives, and reference works; 2008.

339. Roberts A, Gaizauskas R, Hepple M, Demetriou G, Guo Y, Roberts I, et al. Building a semantically annotated corpus of clinical texts. J Biomed Inform 2009;42(5):950–66.

340. Hripcsak G, Rothschild AS. Agreement, the f measure, and reliability in information retrieval. J Am Med Inform Assoc 2005;12(3):296–8.

341. Brants T. Inter-annotator agreement for a German newspaper corpus. In: Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000). Athens, Greece; 2000.

342. Kilicoglu H, Rosemblat G, Fiszman M, Rindflesch TC. Constructing a semantic predication gold standard from the biomedical literature. BMC Bioinformatics 2011;12(1):486.

343. Song D, Chute CG, Tao C. Semantator: annotating clinical narratives with semantic web ontologies. AMIA Summits on Translational Science Proceedings 2012; 20.

344. Roeder C, Jonquet C, Shah NH, Baumgartner WA Jr, Verspoor K, Hunter L: A UIMA wrapper for the NCBO annotator. BMC Bioinformatics 2010;26(14):1800–1.

345. Pratt W, Yetisgen-Yildiz M. A study of biomedical concept identification: MetaMap vs. people. Proc AMIA Symp 2003:529-33.

346. McKight PE, Najab J. Kruskal-Wallis Test. Corsini Encyclopedia of Psychology; 2010.

347. Dinno A. Nonparametric pairwise multiple comparisons in independent groups using Dunn's test. Portland State University Community Health Faculty Publications; 2015.

348. Uzuner Ö, Solti I, Xia F, Cadag E. Community annotation experiment for ground truth generation for the i2b2 medication challenge. J Am Med Inform Assoc 2010;17(5):519-23.

349. Stubbs A, Uzuner Ö. Annotating risk factors for heart disease in clinical narratives for diabetic patients. J Biomed Inform 2015;58:S78-91.

350. Hand D, Mannila H, Smyth P. Principles of Data Mining. Cambridge, MA: MIT Press; 2001

351. Barbar´ a D (Ed.). Bulletin of the Technical Committee on Data Engineering, Special Issue on Mining of Large Databases 1998;21(1).

352. Decker KM, Focardi S. Technology Overview: A Report on Data Mining. Technical Report CSCS TR-95-02, Swiss Scientific Computing Center; 1995.

353. Desikan P, Hsu K, Srivastava J. Data Mining for healthcare Management, SIAM International Conference on Data Mining, Hilton Phoenix East\Mesa, Arizona, USA; 2011.

354. Church KW, Hanks P. Word association norms, mutual information, and lexicography. Comput Linguist 1990;16(1):22-9.

355. Dice LR. Measures of the amount of ecological association between species. Ecology 1945;26:297–302.

356. Dunning T. Accurate methods for the statistics of surprise and coincidence. Comput Linguist 1993;19(1):61–74.

357. Ilayaraja M, Meyyappan T. Mining Medical Data to Identify Frequent Diseases using Apriori Algorithm, In: Proceedings of the 2013 International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME) 2013;(2): 21-22.

358. Patil BM, Joshi RC, Toshniwal D. Association rule for classification of type -2 diabetic patients, Second International Conference on Machine Learning and Computing 2010;(2):9-11.

359. Raskar SS, Thakore D. Text mining and clustering analysis. IJCSNS, 2011;11(6): 203.

360. Dagan I, Lee L, Pereira FC. Similarity-based models of word cooccurrence probabilities. Mach Learn 1999;34(1-3):43-69.

361. Bullinaria JA, Levy JP. Extracting semantic representations from word co-occurrence statistics: A computational study. Behav Res Methods 2007;39(3):510-26.

362. Lind´en K, Piitulainen J. Discovering synonyms and other related words, CompuTerm 2004;(1):63–70.

363. Wartena C, Brussee R, Slakhorst W. Keyword extraction using word co-occurrence. In Database and Expert Systems Applications (DEXA), 2010 Workshop on IEEE 2010;(8): 54-8.

364. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. J R Stat Soc Ser C (Applied Statistics) 1979;28(1):100-8.

365. Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu AY. An efficient k-means clustering algorithm: Analysis and implementation. Pattern Analysis and Machine Intelligence, IEEE Transactions on 2002;24(7):881-92.

366. Moon TK. The expectation-maximization algorithm. Signal processing magazine, IEEE 1996;13(6):47-60.

367. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser A (methodological) 1977;1-38.

368. Agrawal R, Srikant R. Fast algorithms for mining association rules. In Proc. 20th int. conf. very large data bases VLDB 1994;9(1215):487-499.

369. Agrawal R, Imielinski T, Swami A. Mining Association Rules Between Sets of Items in Large Databases, Proc. 1993 ACM SIGMOD, Washington, DC, 1993;5(1):207–16.

370. Srikant R. Fast Algorithms for Mining Association Rules and Sequential Patterns. Ph.D. Thesis, Department of Computer Science, University of Wisconsin, Madison, WI; 1996.

371. Hahsler M, Chelluboina S, Hahsler MM. Package 'arulesViz'. R-project CRAN, 2011.

372. Bremner JD, Southwick S, Brett E, Fontana A, Rosenheck R, Charney DS. Dissociation and posttraumatic stress disorder in Vietnam combat veterans. Am J Psychiatry 1992; 149: 328-32.

373. Gershuny BS, Thayer JF. Relations among psychological trauma, dissociative phenomena, and trauma-related distress: A review and integration. Clin Psychol Rev 1999;19(5):631-57.

374. Gamberger D, Lavrac N, Jovanoski V. High confidence association rules for medical diagnosis, In Proceedings of IDAMAP99 1999;(1): 42-51.

375. Westen D. The scientific legacy of Sigmund Freud: Toward a psychodynamically informed psychological science. Psychol Bull 1998;124(3):333.

376. Tampke AK, Irwin HJ. Dissociative processes and symptoms of posttraumatic stress in Vietnam veterans. J Trauma Stress 1999;12(4):725-38.

377. Feeny NC, Zoellner LA, Fitzgibbons LA, Foa EB. Exploring the roles of emotional numbing, depression, and dissociation in PTSD. J Trauma stress 2000;13(3):489-98.

378. Kardiner A. The traumatic neuroses of war. Washington, DC: National Academies; 1941.

APPENDIX A - Interview questions

1. What are common strategic goals in your research for PTSD?
2. What are they types of data you typically need in order to achieve these goals?
3. Is there common information you are unable to obtain within the domain?
4. What are the kinds of analysis you perform on the data you obtain?
5. Is there analysis you wish to perform but are unable and if so, why?
6. How do you feel data collection could be improved in your department?
7. How do you feel analysis can be improved?
8. What challenges/barriers might be faced in implementing a new terminology system in the workflow of currently used applications?
9. What type of terminology resources do you use to support the applications you develop/use?
10. What do you like about these terminology resources or knowledge bases?
11. What do you not like about these terminology resources or knowledge bases/what are their respective shortcomings?
12. What have been your constraints?
13. Is there anything that I should have asked in order to better understand the research in this domain or could improve research?

APPENDIX B – Selected competency questions

QUESTIONS GATHERED
What are the signs and symptoms of PTSD in the patient clinical notes?
What are the treatments of PTSD in the patient clinical notes?
What finding precedes hallucinations?
What is the relation between treatment concept and symptom concept?
Which treatment is related to this symptom?
Which is the most effective treatment for this symptom?
What is the collection of symptoms that co-occur together?
What is the category for the collection of symptoms that co-occur together?
What is the collection of treatments that co-occur together?
What is the category for the collection of treatments that co-occur together?
Which treatments overlap?
What CAM treatments have been utilized by physicians on patients with headaches?
What are the symptoms of a PTSD patients with comorbid TBI?
Which patients have a depressive episode predated by nightmares?
What treatments target a specific sub-population of cohorts?
Which drug treatments are post-dated by improvement?
What symptoms are complicating other anxiety disorders of patients?
What symptoms co-occur with a PTSD sign or symptom?
What diseases are associated with PTSD symptoms?
Does the patient have the minimum PTSD symptoms as defined by the DSM-IV/V/Other?
What are the most effective treatments for the PTSD patient?
Has the treatment for the patient changed?
What is/are the prior treatment/s for the PTSD patient?
What is the distinction between traumatic simple phobia and PTSD?
What is the clinical phenomenology of prolonged and repeated trauma?
Is the patient compliant in their treatment?
Is there a change in the PTSD symptoms?
Does the patient have the minimum PTSD symptoms as defined by the DSM-IV/V/Other AND have/not have symptoms of TBI?
Does the patient have the minimum PTSD symptoms as defined by the DSM-IV/V/Other AND have/not have symptoms of another anxiety disorder?
What are the protective factors of PTSD?
What is the clinical course of untreated PTSD?
Are there other subtypes of PTSD?

APPENDIX C – Terminology requirements

ID	Requirement	Quality	Priority
REQ1	PTSDO must be able to explicitly commit to specific terminologies, indicating precisely which set of definitions and assumptions are made.	met	essential
REQ2	PTSDO must use concepts from resources that have their own unique identifiers, such as a IRI reference.	met	conditional
REQ3	PTSDO shall reuse existing concepts and terms from available and reputable sources.	met	essential
REQ4	PTSDO shall be maintained to provide exports of all or part of the knowledge in standard XML formats.	met	essential
REQ5	PTSDO shall maintain traceability of the entities represented within the framework via Internationalized Resource Identifier (IRI).	met	essential
REQ6	Concept names must be unique, unambiguous, and provide context to its intended logical and definitional use.	unmet	essential
REQ7	PTSDO must be able to reuse concepts and terms by explicitly extending other terminology resources while adhering to transitive relations.	unmet	conditional
REQ8	PTSDO shall maintain version control procedures.	met	essential
REQ9	Concepts shall be validated prior to input into PTSDO.	unmet	essential
REQ10	User shall not input experimental, trial, or off-label therapeutic intervention concepts.	met	conditional
REQ11	PTSDO shall retain semantic types of imported terms.	met	conditional
REQ12	PTSDO shall assist cTAKES in support of concept extraction.	met	conditional
REQ13	PTSDO shall be compliant with the W3C's Web Ontology Language – OWL specifications.	met	essential
REQ14	PTSDO shall support linking and complementing existing resources that provide information about PTSD concepts.	unmet	conditional
REQ15	PTSDO must be able to support standard data types.	unmet	conditional
REQ16	PTSDO must include features for defining equivalency of classes and properties.	unmet	conditional
REQ17	Each term within PTSDO shall be named using singular nouns.	unmet	conditional
REQ18	User shall not delete retired classes and instead relocate each to the deprecated class.	met	essential
REQ19	PTSDO may support the composition of properties in statements about classes and properties.	unmet	optional
REQ20	PTSDO should include meta-data annotations: a) synonyms (hasExactSynonym); b) description; c) preferred label; d) example of usage; e) creator; f) publish date; g) version information; and h) status.	met	optional
REQ21	PTSDO must include meta-data annotations: a) label; b) definition; c) definition source; d) database cross reference; and e) semantic type.	met	essential

APPENDIX D – Subset from annotation guidelines

1.1 General Annotation Information

In the examples in the guideline below:

- **Bolded text** provides examples of what to annotate
- Underlined text provides examples of what should NOT be annotated
- *Italicized text* provides explanations of examples

Annotation Scope
For all annotations in all classes, the scope is limited to information of concepts represented in this guideline. The definitions have been made as explicit as possible, but some judgment and context may be used in annotating relevant concepts. Many synonyms and modifications exist for each of the concepts, and reviewers should annotate these variations if they meet the strict definition.

2 Concepts to Annotate

2.1 Semantic Types Used in Annotations

Therapeutic or Preventive Procedure Annotation Classes
Annotate concepts or phrases that are of the semantic type Therapeutic or Preventive Procedure. Include any abbreviations or acronyms that represent the Therapeutic or Preventive Procedure types. Any questionable abbreviations or acronyms should be excluded. When in doubt about any concept, phrase, abbreviation, or acronym please exclude it. Include mention of negation in the proper annotation category.

> **Example: Cognitive behavioral therapy** for the **treatment** of pediatric **post traumatic stress disorder**: a review and meta-analysis.
>
> ***Cognitive behavioral therapy*** *= Therapeutic or Preventive Procedure class*
> *post traumatic stress disorder* **=** *Mental or Behavioral Dysfunction class*

Sign or Symptom Annotation Classes
The signs and symptoms for these guidelines are only concerned with any mention of those related to PTSD. Either asserted or negated. In an effort to capture symptoms for PTSD, do not annotate any diagnoses of depression or MDD as a symptom of PTSD.

> Example: This study investigated the interrelationship between **PTSD symptoms** among **people** receiving care at the facility.
>
> ***PTSD Symptoms*** *= Sign or Symptom class*
> *people* **=** *Population Group class*

Example: We highlight why the **traumatic event** happened and why the client was still suffering, resulted in profound emotional distress in session.

traumatic event *= Sign or Symptom class*
emotional distress = Mental or Behavioral Dysfunction class

*While **traumatic event** is of the Event Semantic Type in the UMLS, in the domain of PTSD, we are classifying it as a sign of the disorder.*

Mental or Behavioral Dysfunction Annotation Classes

Example: Multiple trauma **patients** reported more **dissociation** than those that experienced single trauma.

dissociation *= Mental or Behavioral Dysfunction class*

2.3 How to Annotate Classes of Concepts

2.3.1 Create Instance

A user can create a single instance at a time by following the steps: 1. Select a piece of text from the loaded document 2. Right click -> Create Instance (See **Figure 1**) 3. Choose a class from the class list of the popup window (See **Figure 2**) 4. Click "Done" 5. The first time a user creates an instance of a particular class, he/she will be asked to choose a datatype property to store the selected text; also, the system will also ask the user to pick a color to highlight all instances of the selected class.

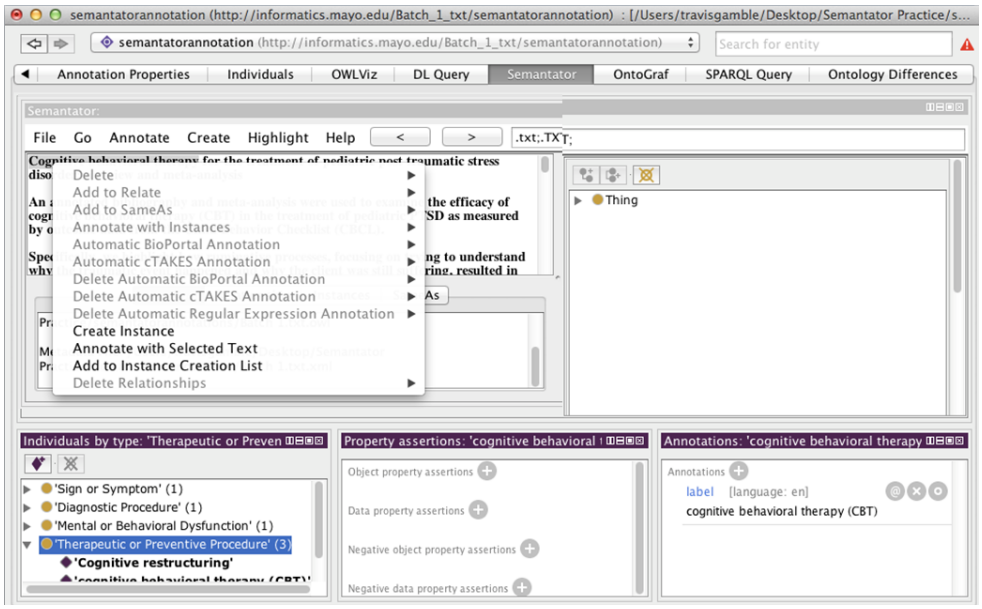


Figure A.1 Create an Instance

APPENDIX E –SOURCE CODE

Source file: GoldStandard.json

```
{
    "id": 25105075,
    "terms": [
        "anxiety",
        "assault",
        "nightmare",
        "CBT",
        "panic",
        "violence",
        "anger management",
        "panic",
        "talk therapy",
        "anxiety",
        "trauma",
        "numbing",
        "anxiety",
        "insomnia",
        "nightmares",
        "case management",
        "anxiety",
        "flashbacks",
        "irritability",
        "temper",
        "depression"
        "pain",
        "psychosocial treatment"
        "trauma",
        "re-experiencing",
        "flashback",
        "anxious",
        "nightmares",
        "insomnia"
        "panic",
        "self-mutilating",
        "fear",
        "hallucination",
        "paranoid thought",
        "cognitive therapy",
        "CT",
        "anger",
        "anxiety",
        "zoloft",
        "deep breathing",
        "anxiety"
    ]
```

APPENDIX F –SOURCE CODE

Source file: features.json

```
{
    "main": "hypervigilance",
    "mainStem": "hypervigilan",
    "grammaticVariations": [
        "hyper-vigilance",
        "hyper vigilance"
    ],
    "grammaticVariationStems": [
        "hyper vigilan",
        "hyper-vigilan"
    ],
    "synonyms": [],
    "synonymStems": []
},
{
    "main": "agoraphobia",
    "mainStem": "agoraphobia",
    "grammaticVariations": [],
    "grammaticVariationStems": [],
    "synonyms": [
        "fear of crowd",
        "crowd",
        "avoid activity",
        "avoid crowd",
        "avoid place",
        "avoid situation",
        "avoid social",
        "avoidance of social",
        "avoidance of social activities",
        "avoid talking",
        "avoid talk",
        "avoid thought",
        "avoid trauma",
        "avoid war film",
        "avoid war",
        "avoid situation"
    ],
    "synonymStems": []
},
{
    "main": "self-blame",
    "mainStem": "blame",
    "grammaticVariations": [],
    "grammaticVariationStems": [],
    "synonyms": [
        "responsible",
        "self critical",
```

```
            "bad person",
            "self-critical",
            "self criticism",
            "self-attacking",
            "self-critical",
            "self-criticism",
            "feel responsible",
            "self blame"
        ],
        "synonymStems": []
    },


    {
        "main": "fear",
        "mainStem": "fear",
        "grammaticVariations": [],
        "grammaticVariationStems": [],
        "synonyms": [
            "frightened",
            "scared",
            "fright",
            "scary",
            "fearful",
            "feared",
            "afraid",
            "horror",
            "apprehension",
            "suspicion",
            "terror"
        ],
        "synonymStems": []
    },


    {
        "main": "flashback",
        "mainStem": "flash",
        "grammaticVariations": [],
        "grammaticVariationStems": [],
        "synonyms": [
            "negative imagery",
            "image",
            "recollection",
            "imagery",
            "persistent recollection"
        ],
        "synonymStems": []
    },
```

APPENDIX G –SOURCE CODE

Source file: core.py

```
#! /usr/bin/env python
# -*- coding: utf-8 -*-

"""
Developed With Python Version 2.7.8

Refer to config.json for setup variables
"""

import sys as S
import csv as C
import os as OS
import json as JSON
import xml.etree.cElementTree as Et
import errno

#
# Utility Methods
#

# Parse JSON Objects to Python Dictionaries
def parseJSONToDicts (path):
	"""
	Parse from strings to dictionaries
	@params:
		path - Required : input path (Str)
	Returns: dicts (List)
	"""
	with open(path, "r") as file:
		dicts = JSON.load(file)
		return dicts

# Drop Duplicate Values in List
def dropDuplicates (list, lowerCase=False):
	"""
	Drop Duplicate Values in List
	@params:
		list - Required : input list (List)
	Returns: new (List)
	"""
```

```
        new = []
        for value in list:
                if lowerCase == True:
                        value = value.lower()
                if value not in new:
        new.append(value)
        return new

# Write Python List of Dictionaries to Directory as JSON File
def writeJSON (data, path="output/output.json"):
        """
        Write JSON output to file.
        @params:
                data - Required : data to write (Dict)
                path - Optional : output path (Str)
        Returns: path (Str)
        """
        with open(path, "w") as file:
                file.write(JSON.dumps(data, indent=4, separators=(',', ': '), sort_keys=False))
                file.close()
        return path

# General File Writing Function
def writeOutput (data, path="output/output.txt"):
        """
        General File Writing Function
        @params:
                data - Required : data to write (Str, Dict, List, Set)
                path - Optional : output path (Str)
        Returns: path (Str)
        """
        with open(path, "w") as file:
                for line in data:
                        file.write(str(line) + "\n")
                file.close()
        return path

# Pipes Output to System Standard Output
def pipeOutput (data):
        """
        Pipes Data to Standard Out
        @params:
                data - Required : data to write (Str, Dict, List, Set)
        """
```

```
        try:
                return sys.stdout.write(data)
        except:
                try:
                        import sys as System
                        return System.stdout.write(data)
        except:
                print "Error: unable to pipe output"

#
# Data Manipulation / Computation Methods
#

def parseFiles (maps, location, attribute):
        output = []
        for dict in maps:
                file = location + str(dict["id"]) + ".xml"
                knownCount = float(len(set(dict["terms"])))
                outputDict = {
                        "id" : dict["id"],
                        "totalFoundTerms" : [],
                        "uniqueFoundTerms" : [],
                        "missingTerms" : [],
                        "knownCount" : knownCount,
                        "foundTotalCount" : 0,
                        "foundUniqueCount" : 0,
                        "missingCount" : 0,
                        "hitRate" : 0,
                        "missRate" : 0
                }
                try:
                        tree = Et.parse(file)
                except:
                        print "Warning: unable to retrieve file ", file
                root = tree.getroot()
                foundTerms = []
                for element in root:
                        for target in element.findall("[@" + attribute + "]"):
                                term = target.get(attribute)
                                for item in dict["terms"]:
                                        if term == item:
                                                foundTerms.append(term)
                outputDict["totalFoundTerms"] = foundTerms
                # Construct Output Dictionary
```

```
                outputDict["uniqueFoundTerms"] = list(set(outputDict["totalFoundTerms"]))
                outputDict["missingTerms"] = list(set(dict["terms"]) –
                                                  set(outputDict["uniqueFoundTerms"]))
                totalFoundTerms = outputDict["totalFoundTerms"]
                uniqueFoundTerms = outputDict["uniqueFoundTerms"]
                missingTerms = outputDict["missingTerms"]
                foundTotalCount = len(totalFoundTerms)
                foundUniqueCount = len(uniqueFoundTerms)
                missingTermsCount = len(missingTerms)
                oddsRatio = None # (foundTotalCount * ) / ()
                if knownCount == 0:
                        totalSuccessRate = None
                        uniqueSuccessRate = None
                else:
                        outputDict["hitRate"] = round((foundUniqueCount / knownCount) * 100,
                        2)
                        outputDict["missRate"] = round((1 - (foundUniqueCount / knownCount))
                        * 100, 4)
                        outputDict["foundTotalCount"] = foundTotalCount
                        outputDict["foundUniqueCount"] = foundUniqueCount
                        outputDict["missingCount"] = missingTermsCount
                        outputDict["oddsRatio"] = oddsRatio
                        output.append(outputDict)
        return output

def aggregateReport (report):
        for dict in report:
                pass

#
# Main Routine
#

if __name__ == "__main__":
        print "\nStatus: report-accuracy.py initialized from commandline\n"
        exitCode = 0
        try:
                # Parse Config & Set Global Variables
                print "Status: configuring"
                config = JSON.loads(S.argv[1])
                version = config["version"]
                gsFile = config["input"][0]["directory"] + config["input"][0]["file"]
                xmlLocation = config["input"][1]["directory"] + config["input"][1]["fileToken"]
                xmlTargetElement = config["input"][1]["element"]
```

```
                outputDirectory = config["output"]["directory"]
                outputJSON = config["output"]["format"]["json"]
                outputCSV = config["output"]["format"]["csv"]
                outputPipe = config["output"]["format"]["pipe"]
                matchOnSubString = config["settings"][0]["value"]
                fileList = []
                print "Status: done\n"
        except:
                print "Error: unable to configure report-accuracy.py; try validating the
                config.json file online at JSONlint\n"
                S.exit(exitCode)
        print "Status: retrieving input files"
        try:
                # Set Data
                gsData = parseJSONToDicts(gsFile)
                print "Status: done\n"
        except:
                print "Error: unable to retrieve data files, ensure directory paths in config.json
                are correct\n"
                S.exit(exitCode)
        print "Status: generating report, this may take a moment . . ."
        try:
                xmlReport = parseFiles(gsData, xmlLocation, xmlTargetElement)
                metaReport = aggregateReport(xmlReport)
                if (outputJSON == True):
                        writeJSON(xmlReport, outputDirectory + "report.json")
                if (outputCSV == True):
                        writeOutput(xmlReport, outputDirectory + "report.csv")
                if (outputPipe == True):
                        pipeOutput(xmlReport)
                if (outputJSON == False and ouputCSV == False and pipeOutput == False):
                        print report, "\n"
                        print "Notification: report not saved! Make sure the output type is set in
                        config.json"
                print "Status: done\n"
        except:
                print "Error: unable to generate report\n"
                S.exit(exitCode)
        exitCode = 1
        S.exit(exitCode)
else:
        pass
```

APPENDIX H –SOURCE CODE

Source file: accuracy-metrics.py

```
#! /usr/bin/env python
# -*- coding: utf-8 -*-

"""
Developed With Python Version 2.7.8

Refer to config.json for setup variables
"""

import sys as S
import csv as C
import os as OS
import os.path as Path
import subprocess as Subprocess
import json as JSON
import xml.etree.cElementTree as Et
import re as Rgx
import errno

#
# Utility Methods
#

# Parse JSON Objects to Python Dictionaries
def parseJSONToDicts (path):
	"""
	Parse from strings to dictionaries
	@params:
		path - Required : input path (Str)
	Returns: dicts (List)
	"""
	with open(path, "r") as file:
		dicts = JSON.load(file)
		return dicts

# Make Directories From Path
def makeDirectories (path):
	"""
	Create directories from path
	@params:
```

```
            path - Required : directory path (Str)
        Returns: path (Str)
        """
        try:
            OS.makedirs(path)
        except:
            try:
                import os as OperatingSystem
                OperatingSystem.makedirs(path)
            except OSError as exception:
                if exception.errno != errno.EEXIST:
                    raise
        return path

# Prompt user input from command line
def getUserInput (valid, prompt):
        """
        Prompts user for and validates input using regular expression
        @params:
            prompt - Required : verbose user prompt (Str)
            valid - Required : regex to validate against (Rgx)
        Returns: dicts (List)
        """
        response = raw_input(prompt)
        if Rgx.match(valid, response):
            return response
        else:
            print "Error: Invalid input"
            getUserInput(valid, prompt)

#
# Main Routine
#

if __name__ == "__main__":
        print "\nStatus: core.py initialized from commandline\n"
        exitCode = 0
        try:
        # Parse Config & Set Global Variables
            print "Status: configuring"
            config = parseJSONToDicts("config.json")
            startupNotification = config["startupNotification"]
            toolConfigs = config["tools"]
            validTools = 0
```

```
        currentTool = 0
        voidTools = []
        option = 1
        print "Status: done\n"
        if startupNotification != False:
                print "Notification:", startupNotification, "\n"
except:
        print "Error: unable to configure core.py; try validating the config.json file online
        at JSONlint\n"
        S.exit(exitCode)
print "Status: verifying tools"
try:
        # Tool Verification Routine
        for tool in toolConfigs:
                if Path.isfile(tool["source"]) == False:
                        print "Warning: Unable to find", tool["name"]
                        voidTools.append(currentTool)
                        currentTool += 1
                else:
                        validTools += 1
                        currentTool += 1
        # Remove Voided Tools From Tool Config Dictionary
        for tool in voidTools:
                del toolConfigs[tool]
        print "Status: done\n"
except:
        print "Error: unable to verify"
        S.exit(exitCode)
print "Input: select tool . . ."
try:
        # User Selection of Tool
        for tool in toolConfigs:
                print "\t[", option, "]", tool["name"]
                option += 1
        selection = int(getUserInput(valid=r"[0-9]{1,2}", prompt="Hint: enter [ n ] to
                select the appropriate tool\nSelection: "))
        print "Status: input received\n"
except:
        print "Error: unable to receive user input"
        S.exit(exitCode)
print "Status: retrieving tool & scaffolding output directory"
try:
        # Scaffold Tool Output Directory
        toolConfig = toolConfigs[selection - 1]
```

```
                toolName = toolConfig["name"]
                toolSource = toolConfig["source"]
                toolVersion = toolConfig["version"]
                toolOutputDirectory = toolConfig["output"]["directory"]
                toolInitCommand = [toolSource, JSON.dumps(toolConfig)]
                makeDirectories(toolOutputDirectory)
                print "Status: done\n"
        except:
                print "Error: unable to setup tool\n"
                S.exit(exitCode)
        print "Status: running", toolName, "- Version:", toolVersion
        # Run Selected Tool
        try:
                returnCode = Subprocess.call(toolInitCommand, close_fds = True)
                if returnCode == 0:
                        raise
                print "Status:", toolName, "ran successfully\n"
        except:
                print "Error: unable to execute", toolName, "\n"
                S.exit(exitCode)
        exitCode = 1
        S.exit(exitCode)
else:
        pass
```

APPENDIX I – Selected association rules

Trauma exposure association rules

Probability	Lift	Rule
82 %	3.52	memory impairment = Existing, relationship difficulty = Existing -> trauma exposure = Existing
78 %	3.19	violent = Existing, cognitive distortion = Existing -> trauma exposure = Existing
75 %	2.85	amnesia = Existing, acute panic = Existing -> trauma exposure = Existing
75 %	2.95	numbness = Existing, dream anxiety = Existing -> trauma exposure = Existing
73 %	2.82	blackout = Existing, acute stress = Existing -> trauma exposure = Existing
72 %	2.65	blackout = Existing, feel inadequate = Existing -> trauma exposure = Existing
72 %	2.75	authority difficulty = Existing, emotional recollection = Existing -> trauma exposure = Existing
71 %	2.66	amnesia = Existing, cognitive distortion = Existing -> trauma exposure = Existing
70 %	6.48	amnesia = Existing, dream anxiety = Existing -> trauma exposure = Existing
70 %	2.62	relationship difficulty = Existing, emotional recollection = Existing -> trauma exposure = Existing
68 %	2.46	blackout = Existing -> trauma exposure = Existing
68 %	2.45	violent = Existing, feel inadequate = Existing -> trauma exposure = Existing
66 %	2.23	memory impairment = Existing, control issue = Existing -> trauma exposure = Existing
66 %	2.12	social dysfunction = Existing, violent = Existing -> trauma exposure = Existing
65 %	2.06	mania = Existing, forget medication = Existing -> trauma exposure = Existing
65 %	12.06	violent = Existing, authority difficulty = Existing -> trauma exposure = Existing
65 %	2.05	academic difficulty = Existing, feel inadequate = Existing -> trauma exposure = Existing
64 %	2.03	authority difficulty = Existing, acute panic = Existing -> trauma exposure = Existing
63 %	1.96	cognitive distortion = Existing, relationship difficulty = Existing -> trauma exposure = Existing
63 %	1.76	self-harm = Existing, violent = Existing -> trauma exposure = Existing
63 %	1.81	violent = Existing, acute panic = Existing -> trauma exposure = Existing
61 %	1.71	academic difficulty = Existing, emotional recollection = Existing -> trauma exposure = Existing
61 %	1.72	self-harm = Existing, cognitive distortion = Existing -> trauma exposure = Existing
58 %	1.32	pain = Existing, cognitive distortion = Existing -> trauma exposure = Existing
58 %	1.37	employment difficulty = Existing, emotional recollection = Existing -> trauma exposure = Existing
58 %	1.32	feel numb = Existing, authority difficulty = Existing -> trauma exposure = Existing
58 %	1.32	headache = Existing, acute stress = Existing -> trauma exposure = Existing
57 %	1.28	alcohol abuse = Existing, emotional recollection = Existing -> trauma exposure = Existing
57 %	1.27	adjustment problem = Existing, arousal reactivity = Existing -> trauma exposure = Existing
57 %	1.26	dysfunctional = Existing, relationship difficulty = Existing -> trauma exposure = Existing
46 %	1.05	self-harm = Existing, forget medication = Existing -> trauma exposure = Existing

Selected association rules

Post-traumatic stress association rules

Probability	Lift	Rule
77 %	5.32	reexperience event = Existing, blunted affect = Existing -> post-traumatic stress = Existing
74 %	7.15	self-blame = Existing, intrusive memory = Existing -> post-traumatic stress = Existing
71 %	2.90	vulnerable = Existing, arousal reactivity = Existing -> post-traumatic stress = Existing
65 %	4.34	feel numb = Existing, emotional abuse = Existing -> post-traumatic stress = Existing
62 %	4.09	feel numb = Existing, emotional recollection = Existing -> post-traumatic stress = Existing
60 %	3.91	hyperarousal = Existing, feel numb = Existing -> post-traumatic stress = Existing
50 %	1.82	lack engagement = Existing, control issue = Existing -> post-traumatic stress = Existing
50 %	2.82	nightmare = Existing, insomnia = Existing -> post-traumatic stress = Existing
48 %	12.59	flashback = Existing, arousal reactivity = Existing -> post-traumatic stress = Existing
48 %	2.53	relationship difficulty = Existing, arousal reactivity = Existing -> post-traumatic stress = Existing
45 %	2.16	self-harm = Existing, arousal reactivity = Existing -> post-traumatic stress = Existing
44 %	1.97	hyperarousal = Existing, depressed = Existing -> post-traumatic stress = Existing
43 %	1.93	intimacy-related stuck point = Existing, violent = Existing -> post-traumatic stress = Existing
43 %	1.87	self-harm = Existing, anxious reaction = Existing -> post-traumatic stress = Existing
42 %	1.70	self-harm = Existing, control issue = Existing -> post-traumatic stress = Existing
42 %	1.72	nightmare = Existing, control issue = Existing -> post-traumatic stress = Existing
41 %	1.60	intimacy-related stuck point = Existing -> post-traumatic stress = Existing
41 %	1.59	self-harm = Existing, dream anxiety = Existing -> post-traumatic stress = Existing
41 %	1.60	hallucination = Existing, depressed = Existing -> post-traumatic stress = Existing
41 %	1.53	hallucination = Existing, control issue = Existing -> post-traumatic stress = Existing
40 %	1.48	dream anxiety = Existing, depressed = Existing -> post-traumatic stress = Existing

Selected association rules

Nightmare association rules

Probability	Lift	Rule
87 %	5.68	academic difficulty = Existing, distress = Existing -> nightmare = Existing
86 %	9.42	memory impairment = Existing, distress = Existing -> nightmare = Existing
80 %	15.32	self-harm = Existing, distress = Existing -> nightmare = Existing
79 %	9.46	distress = Existing, anxious reaction = Existing -> nightmare = Existing
78 %	15.39	emotional abuse = Existing, distress = Existing -> nightmare = Existing
77 %	5.55	distress = Existing, feel inadequate = Existing -> nightmare = Existing
77 %	8.10	physical injury = Existing, distress = Existing -> nightmare = Existing
76 %	10.79	feeling isolated = Existing, distress = Existing -> nightmare = Existing
75 %	14.88	psychotic symptom = Existing, distress = Existing -> nightmare = Existing
75 %	4.65	numbness = Existing, distress = Existing -> nightmare = Existing
74 %	1.02	distress = Existing, forget medication = Existing -> nightmare = Existing
74 %	4.91	dysfunctional = Existing, distress = Existing -> nightmare = Existing
73 %	11.87	adjustment problem = Existing, distress = Existing -> nightmare = Existing
72 %	4.85	impaired adjustment = Existing, distress = Existing -> nightmare = Existing
71 %	2.86	fear = Existing, distress = Existing -> nightmare = Existing
70 %	1.35	distress = Existing -> nightmare = Existing
69 %	4.43	employment difficulty = Existing, distress = Existing -> nightmare = Existing
68 %	4.62	distress = Existing, anhedonia = Existing -> nightmare = Existing
68 %	6.35	distress = Existing, trauma exposure = Existing -> nightmare = Existing
68 %	4.19	alcohol abuse = Existing, distress = Existing -> nightmare = Existing
68 %	4.90	distress = Existing, depressed = Existing -> nightmare = Existing
68 %	4.20	authority difficulty = Existing, distress = Existing -> nightmare = Existing
67 %	4.40	distress = Existing, relationship difficulty = Existing -> nightmare = Existing
67 %	3.97	difficulty communicating = Existing, distress = Existing -> nightmare = Existing
66 %	4.00	blunted affect = Existing, distress = Existing -> nightmare = Existing
65 %	4.01	social dysfunction = Existing, distress = Existing -> nightmare = Existing
63 %	3.65	helplessness = Existing, distress = Existing -> nightmare = Existing
60 %	13.40	substance dependence = Existing, distress = Existing -> nightmare = Existing

APPENDIX I – Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.80	0.70	0.93	356	27	149
SKATE	0.88	0.83	0.95	418	24	87
NCBO Annotator	0.84	0.76	0.93	383	29	122
MetaMap	0.83	0.75	0.92	381	31	124
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.80	0.72	0.90	365	42	140
SKATE	0.82	0.74	0.90	376	40	129
NCBO Annotator	0.80	0.72	0.89	365	47	140
MetaMap	0.82	0.76	0.87	386	56	119
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.68	0.60	0.78	305	88	200
SKATE	0.72	0.64	0.82	322	72	183
NCBO Annotator	0.70	0.61	0.83	308	64	197
MetaMap	0.74	0.70	0.79	354	94	151
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.64	0.56	0.75	285	97	220
SKATE	0.68	0.60	0.78	302	86	203
NCBO Annotator	0.65	0.57	0.74	290	100	215
MetaMap	0.69	0.64	0.75	323	109	182
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.71	0.69	0.73	350	131	155
SKATE	0.74	0.71	0.76	361	113	144
NCBO Annotator	0.76	0.78	0.74	394	135	111
MetaMap	0.76	0.78	0.73	394	144	111

Figure I.1 Exact PubMed symptom matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.87	0.81	0.94	407	27	98
SKATE	0.94	0.92	0.95	467	24	38
NCBO Annotator	0.89	0.84	0.94	426	29	79
MetaMap	0.88	0.84	0.93	425	31	80
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.85	0.80	0.91	405	42	100
SKATE	0.88	0.84	0.91	424	40	81
NCBO Annotator	0.84	0.79	0.89	398	47	107
MetaMap	0.88	0.87	0.89	437	56	68
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.75	0.71	0.80	357	88	148
SKATE	0.81	0.78	0.84	392	72	113
NCBO Annotator	0.82	0.78	0.86	396	64	109
MetaMap	0.84	0.86	0.82	431	94	70
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.69	0.63	0.77	318	97	187
SKATE	0.73	0.67	0.80	338	86	167
NCBO Annotator	0.76	0.73	0.79	371	100	134
MetaMap	0.77	0.76	0.78	382	109	123
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.74	0.73	0.74	371	131	134
SKATE	0.82	0.85	0.79	427	113	78
NCBO Annotator	0.81	0.86	0.76	432	135	73
MetaMap	0.81	0.87	0.75	440	144	65

Figure I.2 Exact and partial PubMed symptom matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.62	0.53	0.75	89	30	79
SKATE	0.65	0.57	0.75	96	32	72
NCBO Annotator	0.62	0.56	0.71	94	39	74
MetaMap	0.65	0.57	0.75	95	31	73
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.36	0.27	0.57	45	34	123
SKATE	0.37	0.28	0.57	47	36	121
NCBO Annotator	0.36	0.27	0.52	46	43	122
MetaMap	0.35	0.27	0.50	46	46	122
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.32	0.27	0.39	45	71	123
SKATE	0.33	0.29	0.39	49	78	119
NCBO Annotator	0.32	0.27	0.38	46	75	122
MetaMap	0.37	0.34	0.40	57	87	111
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.36	0.27	0.53	46	41	122
SKATE	0.36	0.29	0.49	48	49	120
NCBO Annotator	0.31	0.24	0.41	41	58	127
MetaMap	0.36	0.30	0.43	51	67	117
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.43	0.42	0.44	71	91	97
SKATE	0.41	0.43	0.40	73	111	95
NCBO Annotator	0.40	0.43	0.38	72	118	96
MetaMap	0.40	0.43	0.38	73	120	95

Figure I.3 Exact PubMed treatment matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.75	0.71	0.80	119	30	49
SKATE	0.82	0.83	0.81	140	32	28
NCBO Annotator	0.75	0.73	0.76	123	39	45
MetaMap	0.77	0.74	0.80	125	31	43
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.52	0.42	0.68	71	34	97
SKATE	0.57	0.48	0.69	81	36	87
NCBO Annotator	0.52	0.45	0.64	75	43	93
MetaMap	0.53	0.46	0.63	77	46	91
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.64	0.67	0.61	112	71	56
SKATE	0.64	0.70	0.60	117	78	51
NCBO Annotator	0.63	0.66	0.60	111	75	57
MetaMap	0.65	0.73	0.59	123	87	45
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.74	0.73	0.75	122	41	46
SKATE	0.74	0.76	0.72	127	49	41
NCBO Annotator	0.65	0.65	0.65	110	58	58
MetaMap	0.71	0.77	0.66	129	67	39
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.68	0.79	0.59	133	91	35
SKATE	0.66	0.82	0.55	138	111	30
NCBO Annotator	0.64	0.80	0.53	134	118	34
MetaMap	0.65	0.83	0.54	140	120	28

Figure I.4 Exact and partial PubMed treatment matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.78	0.74	0.83	1242	247	431
SKATE	0.82	0.80	0.84	1336	255	337
NCBO Annotator	0.77	0.74	0.80	1233	303	440
MetaMap	0.78	0.76	0.82	1264	284	409
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.43	0.30	0.80	499	125	1174
SKATE	0.46	0.32	0.80	535	132	1138
NCBO Annotator	0.42	0.29	0.76	491	156	1182
MetaMap	0.47	0.33	0.80	549	137	1124
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.63	0.52	0.78	870	238	803
SKATE	0.65	0.55	0.79	918	249	755
NCBO Annotator	0.59	0.49	0.73	824	298	849
MetaMap	0.66	0.58	0.77	972	294	701
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.62	0.51	0.78	856	242	817
SKATE	0.64	0.54	0.78	901	251	772
NCBO Annotator	0.58	0.49	0.73	814	306	859
MetaMap	0.67	0.59	0.79	980	268	693
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.65	0.56	0.76	937	292	736
SKATE	0.67	0.59	0.77	991	301	682
NCBO Annotator	0.67	0.59	0.76	993	313	680
MetaMap	0.68	0.62	0.76	1029	332	644

Figure I.5 Exact psychotherapeutic transcript symptom matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.81	0.78	0.84	1309	247	364
SKATE	0.87	0.89	0.85	1487	255	186
NCBO Annotator	0.84	0.86	0.83	1432	303	241
MetaMap	0.85	0.87	0.84	1449	284	224
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.47	0.33	0.81	550	125	1123
SKATE	0.51	0.37	0.82	611	132	1062
NCBO Annotator	0.45	0.31	0.77	526	156	1147
MetaMap	0.52	0.38	0.82	629	137	1044
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.71	0.62	0.81	1043	238	630
SKATE	0.73	0.66	0.82	1103	249	570
NCBO Annotator	0.72	0.66	0.79	1097	298	576
MetaMap	0.78	0.76	0.81	1268	294	405
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.66	0.57	0.80	948	242	725
SKATE	0.69	0.61	0.80	1017	251	656
NCBO Annotator	0.65	0.57	0.76	948	306	725
MetaMap	0.74	0.69	0.81	1148	268	525
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.71	0.65	0.79	1092	292	581
SKATE	0.75	0.71	0.80	1183	301	490
NCBO Annotator	0.75	0.71	0.79	1192	313	481
MetaMap	0.80	0.80	0.80	1332	332	341

Figure I.6 Exact and partial psychotherapeutic transcript symptom matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.82	0.79	0.85	258	45	70
SKATE	0.83	0.81	0.85	267	49	61
NCBO Annotator	0.79	0.76	0.83	249	52	79
MetaMap	0.80	0.78	0.82	255	56	73
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.43	0.32	0.67	104	52	224
SKATE	0.43	0.33	0.64	107	59	221
NCBO Annotator	0.45	0.35	0.64	114	65	214
MetaMap	0.45	0.35	0.62	115	71	213
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.41	0.31	0.58	103	75	225
SKATE	0.41	0.32	0.56	106	82	222
NCBO Annotator	0.40	0.32	0.52	105	98	223
MetaMap	0.47	0.41	0.54	134	112	194
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.59	0.50	0.73	165	62	163
SKATE	0.60	0.51	0.72	168	65	160
NCBO Annotator	0.58	0.50	0.69	164	73	164
MetaMap	0.58	0.51	0.66	167	85	161
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.69	0.67	0.70	219	92	109
SKATE	0.70	0.69	0.70	225	98	103
NCBO Annotator	0.67	0.67	0.66	221	114	107
MetaMap	0.67	0.71	0.64	232	132	96

Figure I.7 Exact psychotherapeutic transcript treatment matches

Accuracy Metrics

PTSDO						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.87	0.88	0.87	289	45	39
SKATE	0.91	0.95	0.86	312	49	16
NCBO Annotator	0.88	0.91	0.85	298	52	30
MetaMap	0.88	0.93	0.84	304	56	24
PILOTS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.66	0.58	0.78	189	52	139
SKATE	0.67	0.59	0.77	194	59	134
NCBO Annotator	0.64	0.57	0.74	187	65	141
MetaMap	0.67	0.61	0.74	200	71	128
SNOMED-CT						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.62	0.55	0.71	180	75	148
SKATE	0.64	0.59	0.70	194	82	134
NCBO Annotator	0.63	0.60	0.67	198	98	130
MetaMap	0.67	0.67	0.66	220	112	108
NCI Thesaurus						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.84	0.85	0.82	280	62	48
SKATE	0.84	0.87	0.81	284	65	44
NCBO Annotator	0.84	0.88	0.80	289	73	39
MetaMap	0.85	0.94	0.78	308	85	20
UMLS						
Pipeline	**F-measure**	**Recall**	**Precision**	**TPs**	**FPs**	**FNs**
cTAKES	0.83	0.90	0.76	295	92	33
SKATE	0.83	0.93	0.83	305	98	23
NCBO Annotator	0.80	0.90	0.72	296	114	32
MetaMap	0.81	0.95	0.70	313	132	15

Figure I.8 Exact and partial psychotherapeutic transcript treatment matches