ANNOTATING CLINICAL TEXT USING A FAST HEALTHCARE INTEROPERABILITY RESOURCES-BASED DRUG ONTOLOGY IN A NATURAL LANGUAGE PROCESSING SYSTEM

By

Leong Hui Wong

A Capstone Project

Presented to the Department of Medical Informatics and Clinical Epidemiology,

Oregon Health & Science University

School of Medicine

in partial fulfillment of the requirements for the degree

Master of Biomedical Informatics – Clinical Informatics Track

March 2019

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the Master's Capstone Project of

Leong Hui Wong

"Annotating Clinical Text using a Fast Healthcare Interoperability Resources-Based Drug Ontology in a Natural Language Processing System"

has been approved

William Hersh, MD, FACMI, FAMIA, FACP

Table of Contents

Acknowledgements

To:

Bill Hersh, for his enthusiasm and support for this project,

Hongfang Liu and the Mayo Clinic team, for developing the MedXN system and making it free to the world,

Özlem Uzuner and the i2b2 team, for the annotated dataset,

The many unsung developers and contributors of the open-source tools used in this project,

And above all, to my wife Eng Hooi, my parents, and my family, for their love and encouragement,

My deepest gratitude.

Abstract

Objectives

To develop a medication natural language processing (NLP) system that was derived from MedXN and to evaluate its performance in annotating medication name, dose, route, frequency, and duration using the 2009 i2b2 Medication Extraction Challenge dataset.

Methods

RxNorm was serialized into FHIR Medication, MedicationKnowledge, and Substance resources. The new system, MedXN-FHIR, used the FHIR resources to build a drug vocabulary. In addition, the FHIR data models were utilized for ontology-based reasoning. Both MedXN and MedXN-FHIR were used to annotate a test set of 251 discharge summaries. The system annotations were evaluated against i2b2 ground truth annotations for precision, recall, and F_1-measure.

Results

Both systems exhibited good performance with F_1-measures above 0.8 on medication name, dose, and frequency fields. For medication name, MedXN-FHIR produced an increase in precision of 0.0373 over MedXN. Overall, MedXN-FHIR produced a higher precision but lower recall than MedXN for most attributes. MedXN-FHIR performed worst for duration, with a decrease in F_1-measure of 0.204.

Conclusions

A drug terminology was transformed into FHIR resources, and the same resources were reused as a drug ontology to normalize unstructured medication names and attributes through NLP techniques. Further insight into the performance of MedXN-FHIR in assigning RxNorm identifiers would be useful.

Introduction

Unstructured clinical notes and the lack of adoption of modern terminology and exchange standards, such as RxNorm and Fast Healthcare Interoperability Resources (FHIR), continue to pose challenges to healthcare interoperability and the secondary use of electronic health record (EHR) data.

HL7 FHIR

HL7 FHIR is a healthcare IT interoperability standard that was developed by Health Leven Seven International (HL7). Based on modern application programming interface (API) paradigms, FHIR was designed with a focus on extensibility and ease of implementation. The HL7 FHIR standard prescribes API specifications and data models that capture clinical information in a reusable structured format (i.e. “resources”). FHIR is poised to be the de facto standard for healthcare data exchange. Major EHR vendors, Epic and Cerner, and technology company Apple, have adopted FHIR in their products. In its 2019 Interoperability Standards Advisory (ISA), the Office of the National Coordinator for Health Information Technology (ONC) assessed FHIR to be sufficiently mature for various interoperability scenarios (1).

Using FHIR resources as clinical information models

The use of FHIR resources to standardize clinical information is emerging as a feasible and scalable approach. Most notably, Rajkomar et al. and a team from Google normalized and stored 46.9 billion data points that represented 216,221 hospitalizations from EHR systems in FHIR resources (2). These resources were fed into a deep learning model to predict medical events. While the method developed by Rajkomar et al. did not require data harmonization, they reported that the generalizability of their approach would be limited by the lack of data harmonization between hospital sites. Wu et al. developed SemEHR, a clinical text mining system that uses profiled FHIR

Composition resources to store clinical data in the form of discharge summaries (3). SemEHR also uses a number of ontologies, including Drug Ontology (DrON), to normalize clinical data and to provide semantic search capabilities. Interestingly, the SemEHR system also has a mechanism to learn from user feedback to improve its performance.

The integration of FHIR into domain-specific natural language processing (NLP) pipelines has been gaining traction in recent years. Hong et al. developed a clinical data normalization pipeline to extract medication data from clinical notes into FHIR MedicationStatement resources (4). In their work, MedXN was used to extract and normalize medication names. Other medication-related attributes such as strength and dose form were normalized using value sets in a FHIR terminology service. In the oncology domain, Savova et al. developed DeepPhe, a system that combined an NLP pipeline with an oncology-based ontology. DeepPhe was used to extract cancer phenotypes and treatments from unstructured records into FHIR-based models (5). The group also used the Apache Clinical Text Analysis and Knowledge Extraction System (cTAKES) as their NLP pipeline.

Transforming terminologies into FHIR resources

All four groups addressed the challenges of harmonizing unstructured data and achieving semantic consistency either through automated NLP systems that relied on medical vocabularies (3-5), or by avoiding manual harmonization altogether using direct feature learning (2). Even with the advanced machine learning (ML) methods adopted in the latter approach, harmonized data is still required for interoperability between different healthcare providers.

The HL7 FHIR specifications propose the use of FHIR terminology services to address the need for standardized terminology. However, the terminologies themselves present another challenge for health IT implementers. Terminologies that are complex or based on description logics, such as Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT), can be difficult to

implement (6). This complexity is exacerbated by the poor generalizability of automated mapping tools and the need to manage changes in terminology models and content over time (7, 8).

In the face of these challenges, besides using FHIR resources as containers for clinical data as the earlier groups did, one could simplify a terminology by transforming it into FHIR resources (see Figure 1). Proposed by McMurtrie, this approach was demonstrated using the Australian Medicines Terminology (AMT) with the Medserve prototype (9).

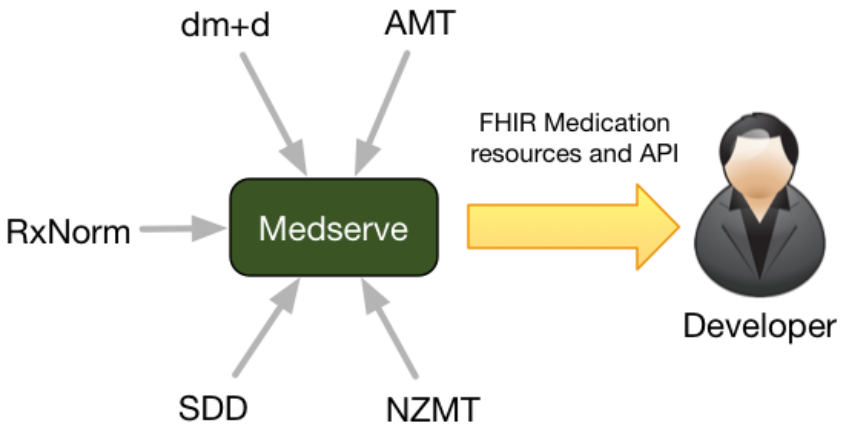

Figure 1: An approach to simplify drug terminologies (reproduced from McMurtrie (9))

Using FHIR resources as an ontology

According to Studer et al., an ontology is defined as "a formal, explicit specification of a shared conceptualization" (10). They further explained the four terms used in their definition as follows:

- "Conceptualization" is an abstract model of ideas or concepts that are grounded in reality.
- "Formal" refers to an ontology that is defined in a machine-readable specification.
- "Explicit" denotes that concepts and relations are defined clearly without ambiguity.
- "Shared" means that the ontology represents consensual knowledge that is accepted by those who use the ontology.

The performance of an NLP system can be improved by integrating an ontology to serve as a knowledge resource and to disambiguate concepts found in the clinical text (11). I hypothesized that FHIR resources, such as Medication resources, can be used as a drug ontology that is simple to implement and light-weight in semantics. Specifically, with a domain ontology, an NLP system can incorporate new inference rules by applying the closed-world assumption (CWA). In CWA, the ontology is assumed to be complete. Using medication brands and ingredients as an example, any associations between brands and ingredients that **are found** in the ontology are assumed to be **true**. Conversely, any associations between brands and ingredients that **are not found** in the ontology are assumed to be **false**. In contrast, in open-world assumption (OWA), an ontology is always assumed to be incomplete and therefore any associations that are not found are not yet known (i.e. rather than assumed to be false).

In this paper, I report a new medication NLP system that uses a FHIR-based drug ontology to normalize unstructured medication name and attributes in clinical text. To my knowledge, this is the first study of its kind where a drug terminology was transformed into FHIR resources, and the same resources were reused as a drug ontology to drive an NLP system.

Research Questions

1. What is the performance of the MedXN system, which was developed using Mayo Clinic data, in annotating clinical text from a different dataset?

2. Is there a difference in performance when a FHIR-based drug ontology is used in MedXN?

Materials and Methods

MedXN

The Medication Extraction and Normalization (MedXN) system was developed by Sohn et al. (the Mayo Clinic team) using 159 randomly-selected clinical notes (12). MedXN uses the Apache Unstructured Information Management Architecture (UIMA) as its underlying NLP platform. UIMA provides a component framework, APIs, and tools for building complex NLP systems (13). Used in IBM Watson and Apache cTAKES, UIMA is also the most frequently utilized framework according to a recent review of clinical information extraction systems (14 – 16).

The Mayo Clinic team designed a series of procedures in the MedXN system. A decomposition-composition strategy was used to extract and rearrange phrases containing medication information following the RxNorm naming convention. Thereafter, MedXN uses the Aho-Corasick algorithm (17) to match the phrases against a medication dictionary, which is derived from RxNorm. MedXN uses three main dictionary files; the ingredient file contains 53,042 phrases for brand names and ingredients while the two medication files contain 175,805 phrases each, in text and in RxNorm concept unique identifiers (RxCUI) formats respectively. When a match is found, MedXN assigns the most appropriate RxCUI to the medication annotation. MedXN was reported to have excellent performance when annotating Mayo Clinic data, with F_1-measures of 0.982 for medication name and between 0.714 and 0.990 for medication attributes.

For this study, I selected the MedXN system because it is open-source. Since its initial publication, other medication information extraction systems have reported similar or superior performance compared to MedXN (18, 19). However, the source codes for those systems were not published. In this study, I used MedXN as a baseline for developing a new system, MedXN-FHIR, and as a comparator for evaluating its performance.

Clinical Corpus

I used the 2009 i2b2 Medication Extraction Challenge dataset, which consists of 1,249 de-identified discharge summaries from Partners Healthcare. Uzuner et al. (the i2b2 team) and the challenge participants annotated 261 summaries to form a set of "ground truth" annotations (20). The i2b2 team split these into 10 and 251 summaries for training and testing respectively. For this study, I used all 261 annotated summaries in the same proportions for training and testing.

To facilitate comparison with MedXN annotations, I wrote a parser to convert the line-token offsets used in the i2b2 annotations into character offsets. For example, the annotation "*aspirin 16:0 16:0*" (starting and ending at line 16 token 1), was converted to "*aspirin::471::491*" (between characters 471 to 491). In the process, I found and corrected nine i2b2 annotations that contain minor errors. These errors include invalid offsets that exceeded the number of tokens in a line, and ending offsets that were smaller in value than the starting offsets.

Drug Vocabulary

I used RxNorm, a terminology standard for all drugs available in the United States (21), as the drug vocabulary for this study. A medication concept in RxNorm consists of discrete information of its ingredient, strength, dose form, and brand. RxNorm also contains synonyms, concept metadata, and mappings to other code systems, such as SNOMED-CT and World Health Organization (WHO) Anatomical Therapeutic Chemical Classification System (ATC) codes. The National Library of Medicine (NLM) distributes RxNorm as one of the source vocabularies of the Unified Medical Language System (UMLS). RxNorm is bundled in UMLS Rich Release Format (RRF) files. Examples of RxNorm concepts and relationships are shown in Figure 2.

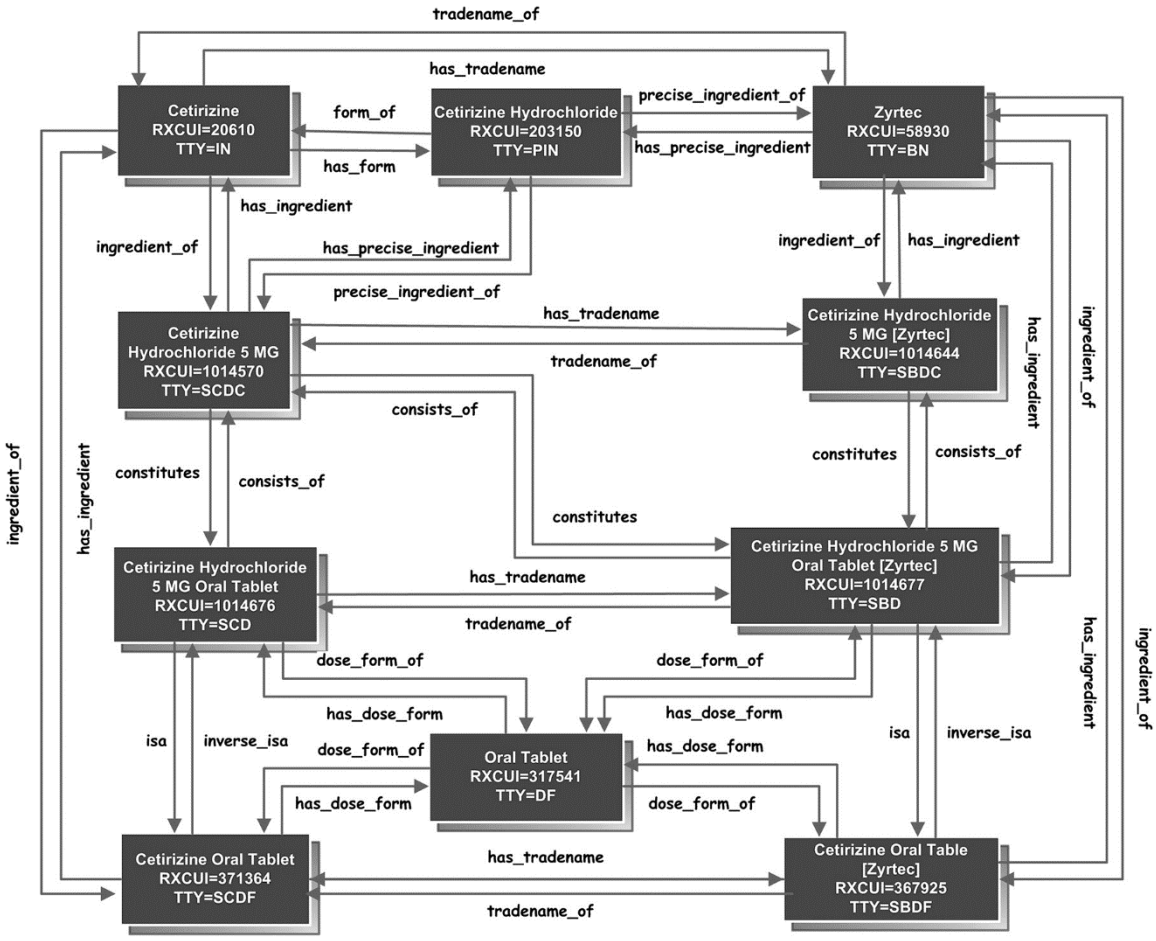

Figure 2: RxNorm model illustrating the concepts and relationships for "*Zyrtec*" (reproduced from Nelson et al. (21))

Development Methodology

To meet the study objectives, I needed to build new features into MedXN, such as FHIR client functions, ontology-based rules, and intensional filters. However, the source code for MedXN had not been updated for the past five years (22). From October to mid-December 2018, I explored and modernized the code base to familiarize myself with the underlying data models (type system) and logic flows. The latter include the algorithm for string matching and decision rules for normalizing annotations and handling special cases. Subsequently, I identified two main development tasks:

1. Serializing RxNorm into a drug ontology that consists of FHIR resources
2. Writing new annotators to query FHIR resources

Using the i2b2 training set, both tasks were performed in parallel from mid-December 2018 to end February 2019 ("the development period").

Serializing RxNorm

RxNorm is available through a FHIR terminology service published by the NLM Value Set Authority Center (VSAC) (23). However, at the time of writing, the VSAC terminology service did not provide any discrete information besides RxCUI and medication name. Therefore, a new procedure is needed to serialize RxNorm concepts, relationships, and attributes information into FHIR resources. I developed a script, RxNormToFHIR, to parse and transform RxNorm in UMLS RRF format into FHIR resources. The data mapping and the resultant FHIR resources are shown in Table 1 and Figure 3 respectively. When FHIR R4 was released mid-way through the project (24), I updated the script to use the new MedicationKnowledge resources to store synonyms for medication names. RxNormToFHIR performs the following procedures:

- Look up each active medication concept against its relationships and attributes.
- Transform each ingredient into a FHIR Substance resource and each concept into a pair of FHIR Medication and MedicationKnowledge resources.
- Represent relationships between RxNorm concepts either as references to other FHIR resources (e.g. "has_ingredient"), or as existing or extension FHIR elements (e.g. "has_tradename").
- Load the resources as bundles into a target FHIR server for remote tests or into JSON files for local tests.

RxNorm Term Type	FHIR Elements
BN (brand name) e.g. *Zyrtec*	Medication.brand (extension)
DF (dose form) e.g. *Oral Tablet*	Medication.form MedicationKnowledge.doseForm
IN/PIN (ingredient/precise ingredient) e.g. *Cetirizine*, *Cetirizine Hydrochloride*	Substance.code Substance.synonym (extension) Medication.ingredient MedicationKnowledge.ingredient
SBD/SCD (branded/clinical drug) e.g. *cetirizine hydrochloride 5 MG Oral Tablet [Zyrtec]*	Medication.code MedicationKnowledge.synonym
SBDC/SCDC (branded/clinical drug component) e.g. *cetirizine hydrochloride 5 MG*	Medication.code MedicationKnowledge.synonym MedicationKnowledge.associatedMedication
SBDF/SCDF (branded/clinical dose form) e.g. *Cetirizine Oral Tablet [Zyrtec]*	Medication.code MedicationKnowledge.synonym MedicationKnowledge.associatedMedication

Table 1: Mapping from RxNorm term types to FHIR elements

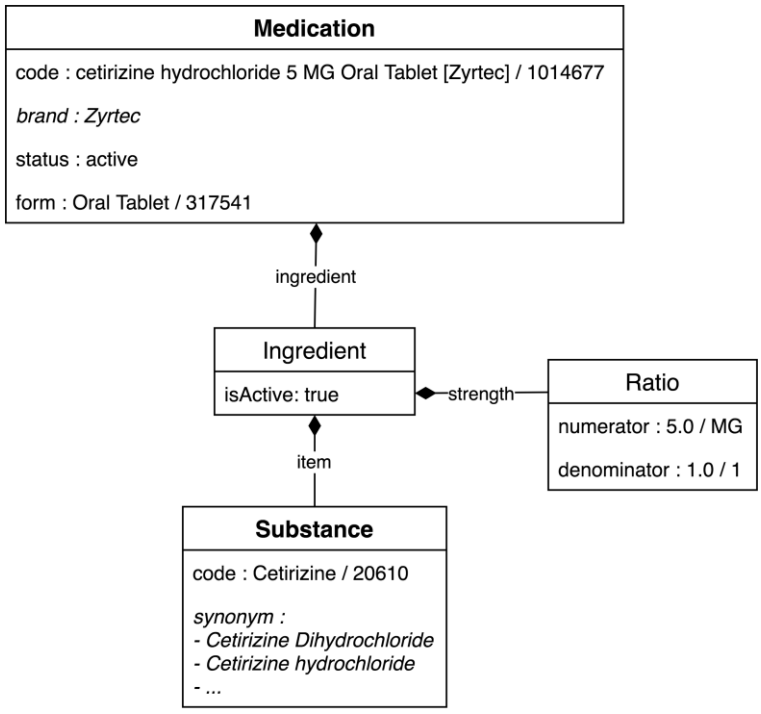

Figure 3: RxNorm concept for "*Zyrtec*" represented as FHIR Medication and Substance resources, with extensions in italics (JSON notation can be found at Appendices A-C)

For the FHIR server, I used the open source HAPI FHIR, which supports FHIR R4 specifications and model structures (25). I configured the FHIR server as a Docker container, which enabled the server to be updated and redeployed rapidly throughout the development period.

Based on the metadata of the RxCUIs found in the MedXN dictionary files, I inferred that the Mayo Clinic team used the January 2012 release of RxNorm. Using RxNormToFHIR, the same RxNorm release was transformed and loaded into the HAPI FHIR server.

Writing new annotators

In UIMA, annotators are software components that implement analysis logic to produce and record annotations (metadata) from document content (13). Since implementing new ontology-base rules would require major modifications to the original annotators, I wrote new annotators for MedXN-FHIR. I also updated component libraries and used a different library for the Aho-Corasick algorithm. The latter task is important because:

1. MedXN annotators depend on pre-built dictionary files. It is not feasible to reuse these files to support structured RxNorm information or to store FHIR resources.
2. Tokenization and matching behavior of the original Aho-Corasick library could not be customized easily, e.g. to ignore overlaps. An example of an overlapping token is "***Provera***" in "*Depo-**Provera***".

Besides an integrated HAPI FHIR client, MedXN-FHIR uses a series of five new FHIR-based annotators. The overall architecture of MedXN-FHIR and a step-wise illustration of its procedures are illustrated in Figure 4 and Figure 5 respectively. The details of each annotation procedure and the differences from MedXN are as follows:

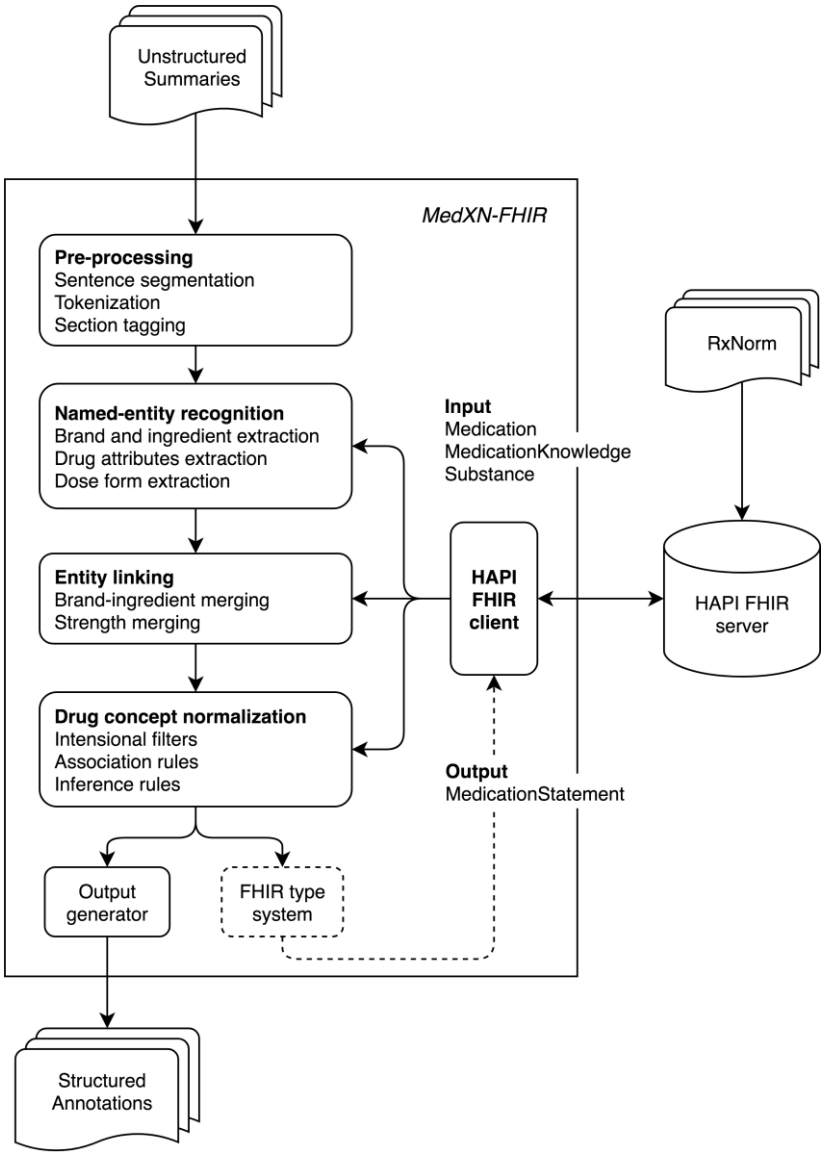

Figure 4: MedXN-FHIR architecture and data flow diagram

Step 1: Pre-processing

MedXN-FHIR uses the same basic NLP pre-processors as MedXN. These pre-processors perform sentence detection, tokenization, chunking, and section tagging.

Step 2: Named-entity recognition (NER)

MedXN-FHIR recognizes drug ingredients and brand names using the same Aho-Corasick string search algorithm. The main differences are:

a. While MedXN reads pre-built dictionary files, MedXN-FHIR queries a FHIR server for Substance and Medication resources to build ingredient and brand vocabularies. The dictionary files in MedXN contain 302 manually curated synonyms, which are not added to MedXN-FHIR.
b. MedXN-FHIR enriches its brand and dose form vocabularies with commonly used synonyms (e.g. "*Eye Drop*" for "*Ophthalmic Drop*"), abbreviations (e.g. "*Tab*" for "*Tablet*"), and plural terms (e.g. "*Suppositories*" for "*Suppository*". MedXN-FHIR generates these synonyms using regular expression rules.
c. While MedXN performs one pass of exact string matching, MedXN-FHIR carries out a second pass using approximate string matching. An edit distance of 1 was used for handling minor misspellings, e.g. "*lorat**i**dine*" instead of "*lorat**a**dine*", while avoiding other similar sounding drug names, e.g. "***r**oxat**i**dine*". The Damerau-Levenshtein metric is used because it considers a transposition, which is a common typographical error, as a single edit (26).

For other medication attributes, MedXN-FHIR uses the same annotator as MedXN, albeit with some modification. Instead of regular expressions, MedXN-FHIR uses RxNorm dose forms that are collated from FHIR Medication resources to build a dose form vocabulary.

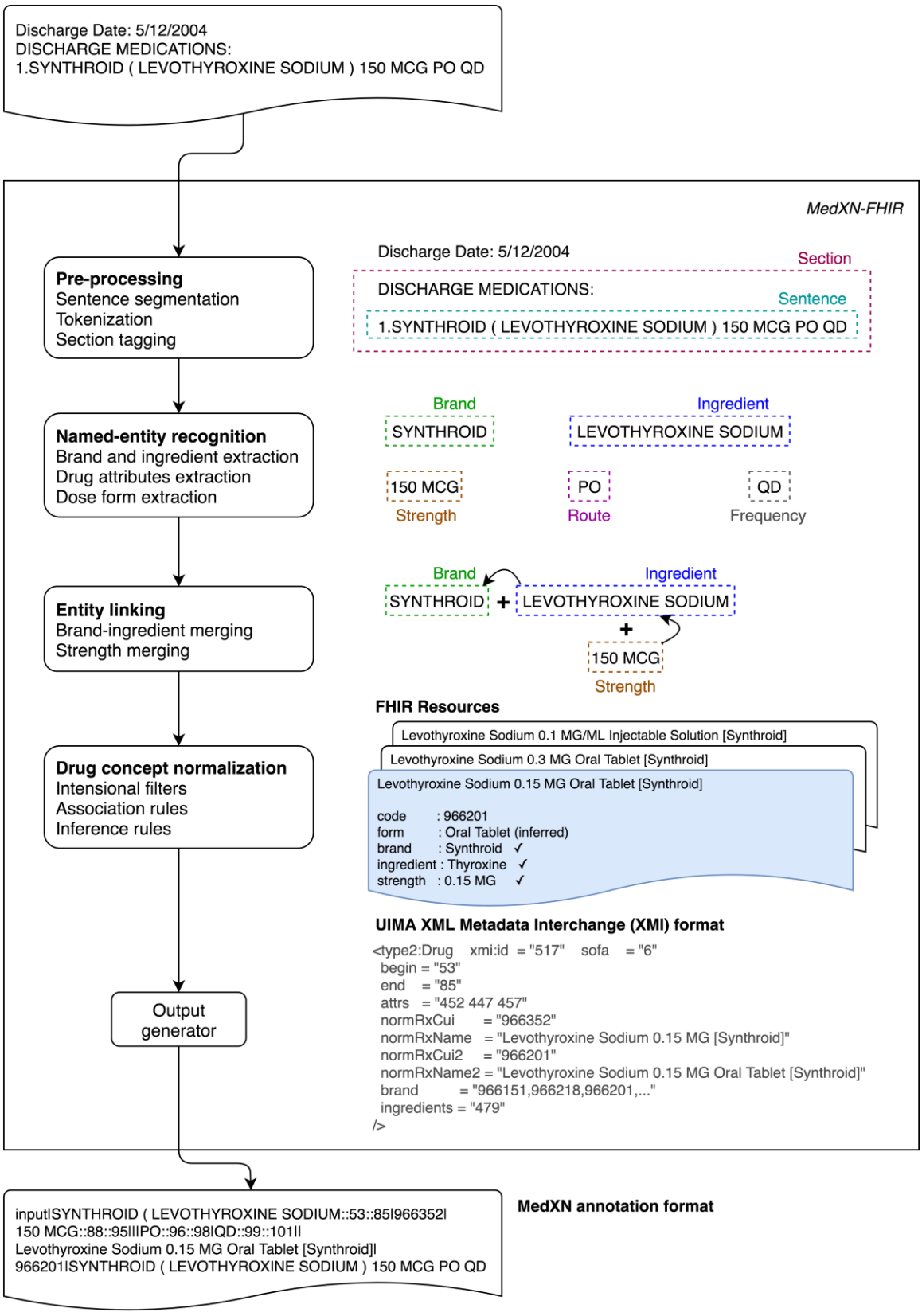

Figure 5: Example of annotation procedures in MedXN-FHIR using "*Synthroid*"

Step 3: Merging annotations

MedXN-FHIR is optimized for annotating medication names. MedXN-FHIR refines annotations by merging adjacent brand and ingredient annotations. MedXN-FHIR uses Medication resources as a drug ontology to validate brand-ingredient merges. To accomplish this, MedXN-FHIR applies CWA using the following rule: for a given brand, the ingredient annotations to be merged must correspond to the ingredients found in Medication resources. For example, "***Zyrtec*** *… Cetirizine*" will be merged but not "***Zyrtec*** *… Ibuprofen*" because there are zero instances of the latter combination in the ontology.

a. MedXN-FHIR uses simple pattern recognition to determine the merge direction, i.e. whether to merge ahead ("***Tylenol*** *… Acetaminophen*") or to merge backwards ("*Acetaminophen …* ***Tylenol***"). MedXN achieves the same, but relies on string matching for parentheses and brackets.
b. A "lookup window" is the distance to look ahead relative to an annotation. To further optimize for medication names, MedXN-FHIR uses a shorter lookup window, which terminates after the mention of a frequency, dose form, or strength. MedXN uses a lookup window of 2 lines of text, which is the same as in the 2009 i2b2 Medication Extraction Challenge.
c. MedXN-FHIR merges each ingredient with adjacent strength annotations. MedXN-FHIR can handle pairs of ingredients and strengths for multi-ingredient drugs. This is shown in greater detail in Appendix D.
d. Lastly, all other attributes, such as route, frequency, and duration are linked to the medication name annotation to form an entry.

Step 4: Drug concept normalization

MedXN-FHIR normalizes medication name annotations through a series of intensionally-defined filters and rules. It attempts to assign two RxCUIs, one based on "stated" facts (*normDrugRxCui*) and another based on inference (*normDrugRxCui2*). Note that the normalization procedure also relies on the assumption that the FHIR resources represent a complete drug ontology (i.e. follows CWA).

Starting with brand and ingredient information extracted in step 2, MedXN-FHIR finds a subset of FHIR Medication and MedicationKnowledge resources. The subset is filtered using intensional rules, i.e. necessary and sufficient conditions defined using attributes in the data model. These rules are used to align the presence or absence of strength, route, and dose form between annotations and FHIR resources. For example, if the annotation does not contain dose form, all FHIR resources in the subset that contain dose form will be excluded. The use of intensional rules could preserve the stability of the normalization process while extensional rules, which are defined using lists of codes, would require continuous review and maintenance (27).

The remaining FHIR resources are then matched based on the populated elements, such as strength value and unit of measure. Only a unique match is assigned as *normDrugRxCui*. If there are no unique matches, MedXN-FHIR attempts to find a less specific RxCUI to increase the rate of RxCUI assignment. From the remaining resources, MedXN-FHIR collects a set of associated concepts. These associated concepts are "parent" concepts that are shared by multiple "child" concepts. For example, "*Cetirizine Oral Tablet [Zyrtec]*" is a parent concept that is shared by "*cetirizine hydrochloride **5 MG** Oral Tablet [Zyrtec]*" and "*cetirizine hydrochloride **10 MG** Oral Tablet [Zyrtec]*".

Lastly, MedXN attempts to assign *normDrugRxCui2* using inference of adjacent contexts. For example, if dose form information is absent, MedXN-FHIR uses route information to infer an appropriate dose form according to RxNorm naming convention. To illustrate, "*Lasix 40mg po"* would be inferred as "*Furosemide 40 MG* ***Oral Tablet*** *[Lasix]"* because there is only one instance of an "oral" dose form among all drugs with the brand "Lasix" and the strength "40 mg" in the ontology.

Another example demonstrating the differences in drug normalization between MedXN and MedXN-FHIR can be found in Appendix E.

Evaluation Design

The performance of MedXN and the new system, MedXN-FHIR, in annotating medication name and attributes were compared using a test set that consists of 251 annotated discharge summaries. In their 2010 article, the i2b2 team measured the performance of twenty competing systems using 3 criteria: phrase vs. token, horizontal vs. vertical, and system-level vs. patient-level (28). These criteria are described as follows:

- A phrase (or exact) match refers to the complete value of a field (e.g. "*sodium chloride*"), while a token (or inexact) match refers to individual words that are delimited by whitespace (e.g. "*sodium*" and "*chloride*").
- Vertical metric refers to performance on individual fields (medication name and attributes) while horizontal metric refers to performance on a medication entry containing all fields spanning two lines of text.
- System-level performance is the average performance over all annotations, while patient-level (or record-level) performance is first measured for individual summaries before being averaged.

For the first criterion, I opted to use only phrase matching because it provides a better measure of normalized medication names. For the second criterion, vertical metric was used because the lookup windows between MedXN and MedXN-FHIR are not equivalent. For the third criterion, both system- and patient-level performance were measured because the former can be affected by summaries that contain disproportionately more medications (29).

All fields that are found in MedXN, MedXN-FHIR, and ground truth annotations (i.e. medication name, dose, route, frequency, and duration) were included for analysis. Both systems generated separate annotations for dose (e.g. "*2 tablets*") and strength (e.g. "*40 mg*") whereas the i2b2

annotators considered them to be the same. To avoid omitting data, I combined dose and strength annotations for each medication entry prior to analysis. I excluded additional fields that are only found in the ground truth annotations, such as medication reason, temporal marker, and certainty. Since the i2b2 ground truth annotations did not contain RxCUIs, the performance of both systems on RxCUI assignment were not compared.

For this study, annotations are classified as correct (true positive) if the offsets are fully bounded by the offsets of the corresponding ground truth annotation. Correct annotations are further filtered to remove overlapping annotations to avoid overestimating the number of true positives. All other annotations are classified as incorrect (false positive). Three performance measures, i.e. precision, recall, and F_1-measure, were calculated and averaged for individual summaries (patient-level) and across all summaries (system-level). If the system fails to annotate any field in the entire summary, as compared to ground truth annotations, a score of zero was imputed.

For patient-level results, each summary constituted a sample and two annotations, one from each system, formed a pair of observations. Using this matching, I performed a paired sample T-test with p = 0.05 using R version 3.5.2 (30). No statistical tests were performed for system-level results.

The performance measures are calculated using the following equations:

$$Precision, P = \frac{\#Retrieved\ phrases\ that\ are\ also\ relevant}{\#Retrieved\ phrases}$$

$$Recall, R = \frac{\#Retrieved\ phrases\ that\ are\ also\ relevant}{\#Relevant\ phrases}$$

$$F - measure = \frac{(1+\alpha)RP}{\alpha R + P}, where\ \alpha = 1$$

Results

The new MedXN-FHIR system was developed over 2.5 months. MedXN-FHIR was released as open-source (https://github.com/leonghui/MedXN-FHIR/) under the same Apache 2.0 license as MedXN. An example of a text annotation using MedXN is shown in Figure 6.

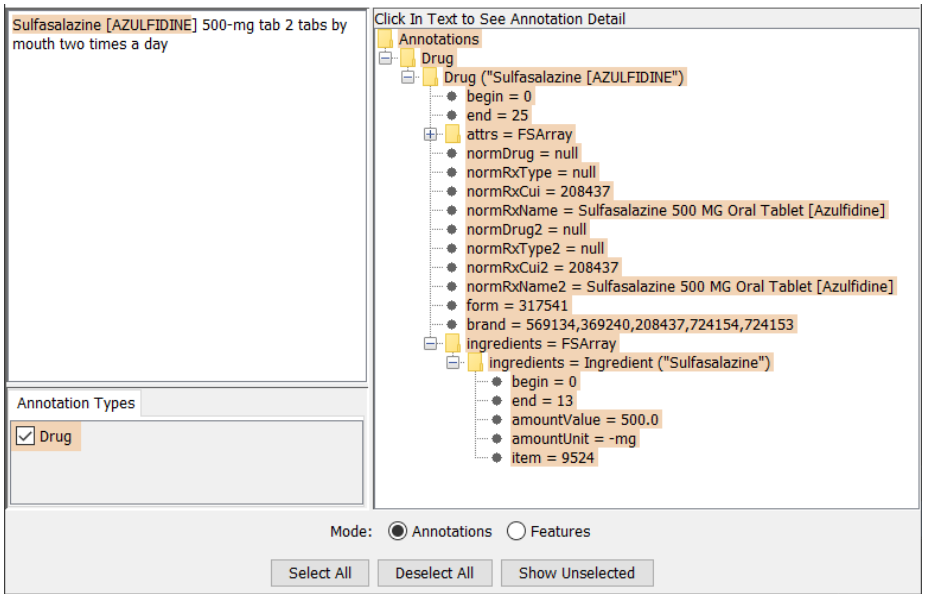

Figure 6: MedXN-FHIR annotation output for "*Azulfidine*"

A transformation script, RxNormToFHIR, was developed over the same period. RxNormToFHIR was also released as open-source (https://github.com/leonghui/RxNormToFHIR) under the more permissive MIT license. Using the January 2012 release, MedXN-FHIR dynamically generated a drug vocabulary that consists of:

- 14,521 ingredient keywords (from 5,010 IN and 1,569 PIN concepts)
- 49,409 brand keywords (from 15,537 BN concepts)
- 307 dose form keywords (from 100 DF concepts)
- 91,024 pairs of Medication and MedicationKnowledge resources (from the same number of SCD, SBD, SCDC, SBDC, SCDF, and SBDF concepts)

All 251 discharge summaries in the test set were processed using MedXN and MedXN-FHIR to obtain medication name, dose, route, frequency, and duration annotations. An overview of the annotation results are tabulated in Table 2.

Source	**Ground Truth**		**MedXN**		**MedXN-FHIR**	
Field	Total	Average	Total	Average	Total	Average
Medication Name	8490	33.8	7079	28.2	6988	27.8
Dose	4381	17.5	3667	14.6	3364	13.4
Route	3303	13.2	2585	10.3	2417	9.6
Frequency	3958	15.8	3595	14.3	3324	13.2
Duration	508	2.0	144	0.6	48	0.2

Table 2: Distribution of fields in the i2b2 ground truth, MedXN, and MedXN-FHIR annotations (overlapping annotations removed, number of discharge summaries = 251)

MedXN and MedXN-FHIR annotations were evaluated against i2b2 ground truth annotations using three performance measures; precision, recall, and F_1-measure. The results are shown in Table 3.

Field	Level	MedXN			MedXN-FHIR		
		P	**R**	**F**	**P**	**R**	**F**
Medication	System	0.8005	0.8359	0.8178	0.8482	0.8231	0.8354
Name	Patient	0.8078	0.8294	0.8116	0.8451	0.8166	0.8231
Dose	System	0.9291	0.8370	0.8806	0.9570	0.7679	0.8521
	Patient	0.9115	0.8367	0.8649	0.9325	0.7747	0.8371
Route	System	0.9638	0.7826	0.8638	0.9821	0.7318	0.8387
	Patient	0.9131†	0.7527	0.8095	0.9080†	0.6935	0.7715
Frequency	System	0.9127	0.9083	0.9105	0.9745	0.8398	0.9022
	Patient	0.9200	0.9171	0.9126	0.9667	0.8483	0.8971
Duration	System	0.8571	0.2835	0.4260	0.8889	0.0945	0.1708
	Patient	0.4264	0.2746	0.3169	0.1957	0.0879	0.1128

† All patient-level measures, except precision for the field "route", were significantly different ($p < 0.05$). No statistical tests were performed for system-level measures.

Table 3: Phrase-level vertical performance of MedXN and MedXN-FHIR using the i2b2 test set, where P = Precision, R = Recall, F = F_1-measure

Differences in performance are quoted using patient-level measures, with the 95% confidence interval enclosed in parentheses. Both MedXN and MedXN-FHIR exhibited good performance for medication name, dose, and frequency fields with F_1-measures above 0.8. For medication name, MedXN-FHIR produced an increase of 0.0373 in precision (0.0284 – 0.0462), a decrease of 0.0128 in recall (0.0067 – 0.0189), and an increase in F_1-measure of 0.0115 (0.0058 - 0.0173). Except for patient-level duration and route, MedXN-FHIR consistently produced a higher precision than MedXN. However, this improvement in precision is counterbalanced by lower recall across all medication attributes. MedXN-FHIR performed worst for the duration field, with a decrease in F_1-measure of 0.204 (0.157 – 0.252).

Discussion

While MedXN-FHIR shares some features with MedXN (source vocabulary, decomposition strategy, attribute annotator), MedXN-FHIR uses ontology-based reasoning for merging and normalizing drug annotations to RxNorm identifiers. Unfortunately, the performance of MedXN-FHIR in assigning RxCUI could not be evaluated in this study due to the lack of RxCUI annotations in the i2b2 dataset.

For medication name, MedXN-FHIR has a higher precision, which could be an outcome of the brand-ingredient merging procedure. The procedure may have helped to reduce the number of overlapping annotations and therefore reduce the number of false positives. On the other hand, MedXN's higher recall could be due to the addition of manually curated synonyms (less false negatives). Medication attributes are linked to a medication name in an entry, i.e. the number of medication attributes retrieved increases with the number of medication names retrieved. Therefore, the tradeoffs that are observed between precision and recall for medication attributes in MedXN-FHIR could be partially attributed to the same factors affecting medication name.

Comparing across system- and patient-level performance, both systems exhibited similar stability except for the duration field, where MedXN-FHIR performed poorly. This result could be due to the difference in the number of duration fields that were retrieved by the systems. Referring to Table 2, MedXN-FHIR had one-third the number of duration annotations as compared to MedXN. Due to the shorter lookup window in MedXN-FHIR, the system was unable to capture the duration terms that extended across drug tokens or beyond one line of text. It is also worthwhile to note that the duration field posed a similar challenge for MedXN in its initial study (F_1-measure = 0.645) (12) and for the top 10 participants of the 2009 i2b2 Medication Extraction Challenge (F_1-measure

between 0.180 and 0.525). The i2b2 team concluded that the duration field was difficult to annotate because of the greater length and variability of the content (28).

Overall, this is a positive result for MedXN-FHIR because its performance was not compromised by the replacement of its main annotators. Furthermore, both MedXN and MedXN-FHIR are comparable to the top 10 i2b2 participants in the phrase-level vertical category (excluding the "medication reason" field). Even though MedXN was built using a significantly different dataset (31), the system also maintained a relatively good performance.

Compared to MedXN, MedXN-FHIR has the added advantage of a FHIR-based ontology. With the flexibility of FHIR data models, the drug ontology can be updated with newer versions of RxNorm to handle new medications and other changes in the terminology. Alternatively, the drug ontology can be built using older versions of RxNorm. This flexibility is useful to account for semantic shifts when annotating unstructured notes from different time periods. Unlike its predecessor, MedXN-FHIR can seamlessly switch between different FHIR servers, each containing a different set of resources. Moreover, since the merging and normalization rules in MedXN-FHIR are defined intensionally using attributes from FHIR data models, only minor adjustments will be needed for MedXN-FHIR to work with other drug terminologies. As demonstrated using RxNormToFHIR, the universality of FHIR models can be exploited to transform local terminologies, such as those used in health systems, and other national terminologies into FHIR resources.

These advantages make MedXN-FHIR suitable for a number of use cases. MedXN-FHIR can be used as part of an audit system to monitor clinical data quality for both structured and unstructured data. For example, MedXN-FHIR can help validate the accuracy of existing mappings and assess the completeness of medication records. Additionally, with its inference feature, MedXN-FHIR can be used to suggest RxCUIs that most closely represent drug products that are

available in the pharmacy. This feature can help reduce the initial transcribing effort for pharmacists who are working with free-text prescriptions. In terms of integration patterns, MedXN-FHIR can be modularized within NLP pipelines, as in NLP2FHIR (4), or scaled up to handle enterprise-level loads using Distributed UIMA Cluster Computing (DUCC).

Lastly, this study demonstrated the feasibility of transforming a drug terminology into FHIR resources and reusing the resources as a drug ontology for a medication NLP system. With the extensibility and ease of implementation of FHIR, this approach could be beneficial in distributing terminology standards. National drug terminologies, such as AMT, Dictionary of Medicines and Devices (dm+d), and Singapore Drug Dictionary (SDD), are generally developed and mandated as de jure standards for healthcare interoperability. It follows that a national drug terminology and an accompanying NLP component can be distributed to heath IT implementers in a format that can be loaded into FHIR-enabled systems readily. FHIR resources can be used without the need for terminology services or description logic reasoners. The potential benefits are two-fold:

1. Reducing the burden of health IT implementers in maintaining mappings from local to national terminologies. RxNorm distributed as FHIR Medication resources could be used as a basis for a drug formulary within EHR and pharmacy systems. No further mapping is necessary as the concept identifiers are already in RxCUI format. For new implementation sites, a “prescribable” subset of FHIR Medication resources with drug names from RxTerms would be especially useful.
2. Minimizing the creation and consumption of unstructured data at source (i.e. EHR and pharmacy systems). FHIR resources are structured and can be used to support clinical data entry. Additionally, NLP can be applied to normalize new unstructured data. Hence, a “virtuous circle” of terminology standardization can be established.

While ML techniques offer us a glimpse of a solution to the "grand challenges" of healthcare informatics (32), I believe that the pursuit of standardization can complement the efforts of ML practitioners in expanding their field into actual care settings. By harmonizing the semantics of clinical data across institutions, the generalizability of ML models can be improved.

Limitations

The performance of MedXN-FHIR in assigning RxCUIs was not evaluated. Future studies could incorporate a panel of independent reviewers to evaluate the RxCUI annotations using the same performance measures. The evaluation protocol should account for additional inferences made by human reviewers that would unintentionally increase false negatives, as observed by the Mayo Clinic team in their work on MedXN (12).

One of the causes of lower precision for MedXN is the presence of HTML entities ("*&*") that were wrongly annotated as the ingredient "*amp"* (adenosine monophosphate). A simple modification to exclude these textual artifacts would have increased MedXN's precision for medication names by 0.05. Clinical corpora used in the future should be cleansed to prevent such artifacts from affecting the analysis.

The current iteration of MedXN-FHIR only validates the merging of brand and ingredient annotations. Future versions of MedXN-FHIR could use the drug ontology to validate the merging of strengths to ingredients. Another component that would be useful in matching strength information is a conversion library that could handle the units of measure used for medications, such as weight/weight concentrations, percentage strengths, and conversion factors for specific drugs (e.g. penicillin and insulin).

MedXN-FHIR's poor performance for the duration field could be addressed by increasing its lookup window. Alternatively, new medication attribute annotators should be developed using a

more structured approach such as those applied by McTaggart et al. (33), with output that conform to the FHIR specifications (i.e. Dosage structure and Timing datatype).

For the analysis in this study, the MedXN annotation output format was reused in MedXN-FHIR without modification. Future developers could enhance MedXN-FHIR to generate the annotation results as FHIR MedicationStatement resources as depicted in Figure 4 (in dashed lines). The integrated FHIR client could be used to send the resources to a FHIR-enabled recipient system (e.g. EHR system, pharmacy system, or Apple HealthKit) as part of an end-to-end workflow.

Conclusion

I developed a medication NLP system that was derived from MedXN. The new system, termed MedXN-FHIR, uses FHIR Medication, MedicationKnowledge, and Substance resources as a drug ontology. In addition to using the resources to build a drug vocabulary, I extended the use of FHIR data models for ontology-based reasoning in MedXN-FHIR.

I evaluated the performance of MedXN and MedXN-FHIR in annotating medication name, dose, route, frequency, and duration fields using the 2009 i2b2 Medication Extraction Challenge dataset. MedXN-FHIR had higher precision, but poorer recall than MedXN. MedXN-FHIR also produced F_1-measures higher than 0.8 for medication name, dose, and frequency.

This study demonstrated the feasibility of transforming a drug terminology into FHIR resources, and reusing the same resources as a drug ontology to normalize unstructured medication names and attributes through NLP techniques.

Appendix

A. FHIR Medication representation of "*Zyrtec*" in JSON notation


```
{
    "resourceType": "Medication",
    "id": "rxNorm-1014677",
    "meta": {
        "versionId": "1",
        "lastUpdated": "2019-03-03T07:01:53.805+00:00"
    },
    "text": {
        "status": "generated",
        "div": "<div xmlns=\"http://www.w3.org/1999/xhtml\"><div
class=\"hapiHeaderText\">cetirizine hydrochloride 5 MG Oral Tablet
[Zyrtec]</div></div>"
    },
    "extension": [
        {
            "url": "http://localhost:8080/fhir/StructureDefinition/brand",
            "valueString": "Zyrtec"
        }
    ],
    "code": {
        "coding": [
            {
                "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
                "code": "1014677",
                "display": "cetirizine hydrochloride 5 MG Oral Tablet [Zyrtec]"
            }
        ]
    },
    "status": "active",
    "form": {
        "coding": [
            {
                "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
                "code": "317541",
                "display": "Oral Tablet"
            }
        ]
    },
    "ingredient": [
        {
            "itemReference": {
                "reference": "Substance/rxNorm-20610"
            },
            "isActive": true,
            "strength": {
                "numerator": {
                    "value": 5,
                    "unit": "MG"
                },
                "denominator": {
                    "value": 1,
                    "unit": "1"
                }
            }
        }
    ]
}
```

B. FHIR MedicationKnowledge representation of "*Zyrtec*" in JSON notation


```
{
    "resourceType": "MedicationKnowledge",
    "id": "rxNorm-1014677",
    "meta": {
        "versionId": "1",
        "lastUpdated": "2019-03-03T10:16:36.710+00:00"
    },
    "code": {
        "coding": [
            {
                "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
                "code": "1014677",
                "display": "cetirizine hydrochloride 5 MG Oral Tablet [Zyrtec]"
            }
        ]
    },
    "status": "active",
    "doseForm": {
        "coding": [
            {
                "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
                "code": "317541",
                "display": "Oral Tablet"
            }
        ]
    },
    "synonym": [
        "Zyrtec 5 MG Oral Tablet",
        "Zyrtec (cetirizine dihydrochloride 5 MG) Oral Tablet"
    ],
    "associatedMedication": [
        {
            "reference": "Medication/rxNorm-367925"
        },
        {
            "reference": "Medication/rxNorm-1014570"
        },
        {
            "reference": "Medication/rxNorm-1014644"
        }
    ],
    "ingredient": [
        {
            "itemReference": {
                "reference": "Substance/rxNorm-20610"
            },
            "isActive": true,
            "strength": {
                "numerator": {
                    "value": 5,
                    "unit": "MG"
                },
                "denominator": {
                    "value": 1,
                    "unit": "1"
                }
            }
        }
    ]
}
```

C. FHIR Substance representation of "*Cetirizine*" in JSON notation


```
{
    "resourceType": "Substance",
    "id": "rxNorm-20610",
    "meta": {
        "versionId": "1",
        "lastUpdated": "2019-03-03T04:16:19.443+00:00"
    },
    "extension": [
        {
            "url": "http://localhost:8080/fhir/StructureDefinition/synonym",
            "valueString": "Acetic acid, (2-(4-((4-chlorophenyl)phenylmethyl)-1-
piperazinyl)ethoxy)-"
        },
        {
            "url": "http://localhost:8080/fhir/StructureDefinition/synonym",
            "valueString": "Cetirizine Dihydrochloride"
        },
        {
            "url": "http://localhost:8080/fhir/StructureDefinition/synonym",
            "valueString": "Cetirizine hydrochloride"
        },
        {
            "url": "http://localhost:8080/fhir/StructureDefinition/synonym",
            "valueString": "(2-(4-((4-Chlorophenyl)phenylmethyl)-1-
piperazinyl)ethoxy)acetic Acid"
        }
    ],
    "status": "active",
    "code": {
        "coding": [
            {
                "system": "http://www.nlm.nih.gov/research/umls/rxnorm",
                "code": "20610",
                "display": "Cetirizine"
            }
        ]
    }
}
```

D. Annotation for multi-ingredient drugs

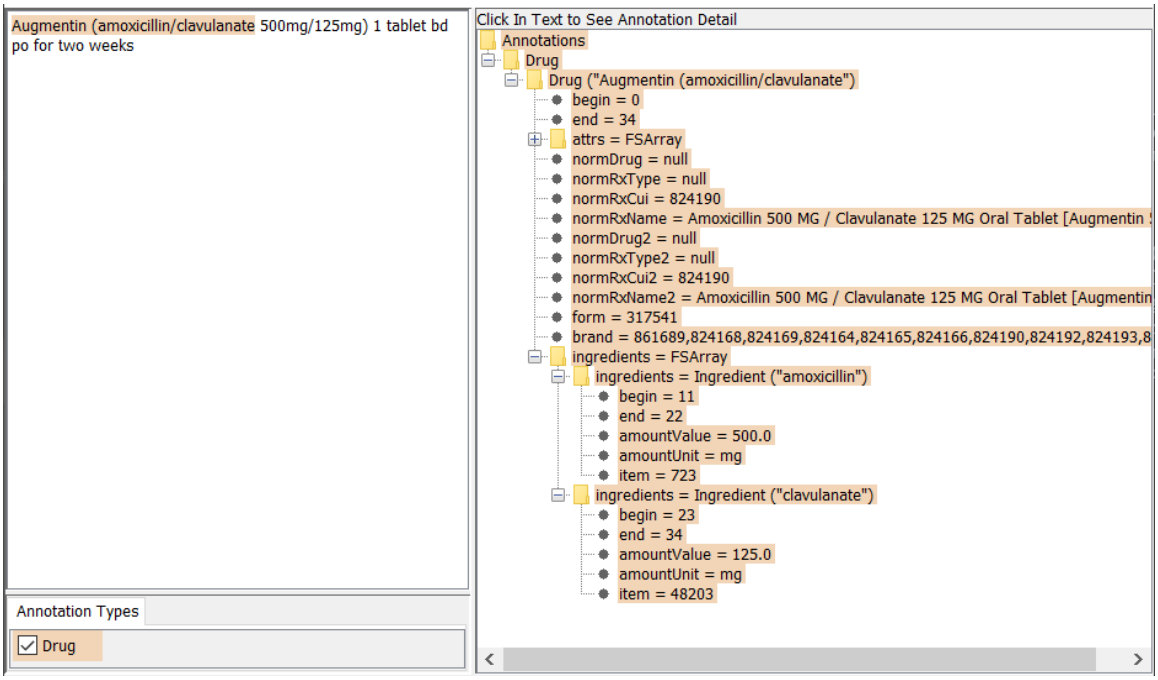

Figure 7: MedXN-FHIR annotation output for a multi-ingredient drug "*Augmentin*"

Figure 7 shows an outcome of MedXN-FHIR's capability in merging the ingredients "*amoxicillin*" and "*clavulanate*" to the brand "*Augmentin*". MedXN-FHIR uses matching Medication resources to validate the merge. The ingredients to be merged must be found among the ingredients of Medication resources with the brand "*Augmentin*". Next, strength attributes were linked to ingredients in the same order ("*500mg*" to "*amoxicillin* and "*125mg*" to "*clavulanate*").

E. Comparison in drug normalization

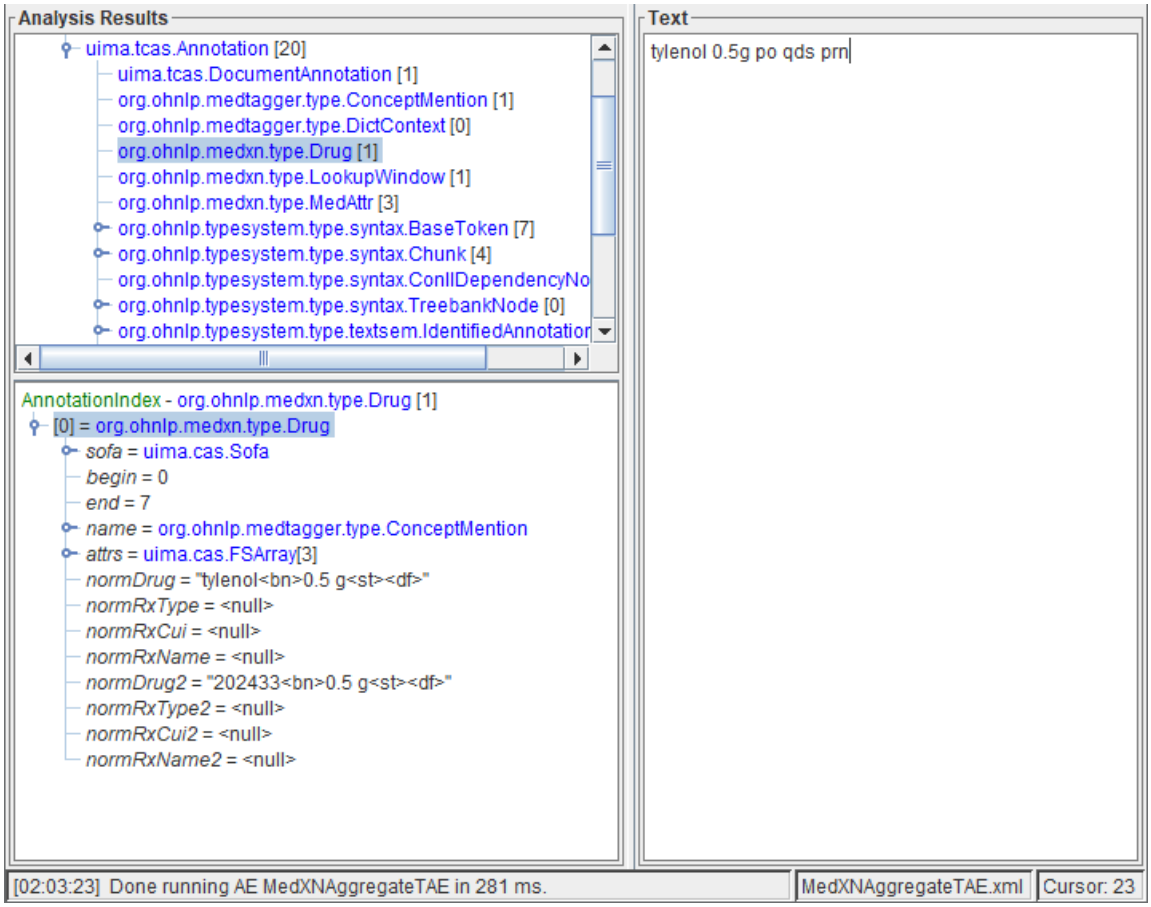

Figure 8: MedXN annotation output for "*Tylenol*"

In Figure 8, note that MedXN was able to recognize "*tylenol*" and "*0.5 g*" as brand and strength respectively. However, MedXN did not assign this annotation with an RxCUI because the phrase "*tylenol 0.5g*" was not found in its dictionary files.

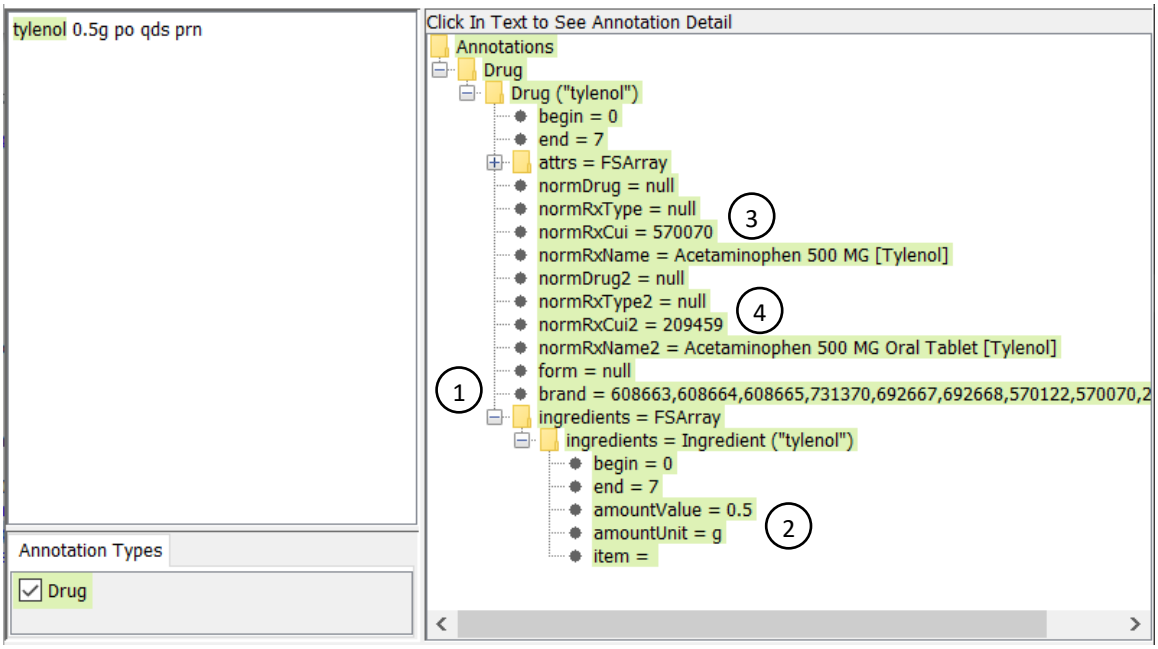

Figure 9: MedXN-FHIR annotation output for "*Tylenol*"

In Figure 9, MedXN-FHIR assigned an appropriate RxCUI using a different set of extraction and normalization rules:

1. **(Named-entity recognition)** MedXN-FHIR collated all Medication resources with the brand "*tylenol*" and stored the RxCUIs in the *Drug.brand* annotation.
2. **(Strength extraction)** MedXN-FHIR extracted the strength "*0.5 g*" as discrete value and unit attributes. To facilitate matching, the strength would be converted to *"500 mg"*.
3. **(Drug concept normalization)** MedXN-FHIR compared each collated Medication resource against the extracted strength information. Since a unique match was found, the system assigned "570070", the identifier of the matched resource, as *normRxCui*.
4. **(Inference)** After normalizing the route "*po*" to "*oral*", MedXN-FHIR checked each Medication resource for an oral dose form. Again, since a unique match was found, the system assigned "209459", the identifier of the inferred resource, as *normRxCui2*.

References

1. Office of the National Coordinator for Health IT. 2019 Interoperability Standards Advisory Reference Edition [Online]. Washington, DC: U.S. Department of Health and Human Services; [updated 2019 January 14; cited 2019 February 26]. Available from: https://www.healthit.gov/isa/sites/isa/files/inline-files/2019ISAReferenceEdition.pdf

2. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med. 2018;1(1):18.

3. Wu H, Toti G, Morley KI, Ibrahim ZM, Folarin A, Jackson R, et al. SemEHR: A general-purpose semantic search system to surface semantic data from clinical notes for tailored care, trial recruitment, and clinical research. J Am Med Inform Assoc. 2018;25(5):530-7.

4. Hong N, Wen A, Shen F, Sohn S, Liu S, Liu H, et al. Integrating structured and unstructured EHR data using an FHIR-based type system: a case study with medication data. AMIA Jt Summits Transl Sci Proc. 2018;2017:74-83.

5. Savova GK, Tseytlin E, Finan S, Castine M, Miller T, Medvedeva O, et al. DeepPhe: a natural language processing system for extracting cancer phenotypes from clinical records. Cancer Res. 2017;77(21):e115-8.

6. Lee D, Cornet R, Lau F, de Keizer N. A survey of SNOMED CT implementations. J Biomed Inform. 2013;46(1):87-96.

7. Stanfill MH, Williams M, Fenton SH, Jenders RA, Hersh WR. A systematic literature review of automated clinical coding and classification systems. J Am Med Inform Assoc. 2010;17(6):646-51.

8. Saitwal H, Qing D, Jones S, Bernstam EV, Chute CG, Johnson TR. Cross-terminology mapping challenges: a demonstration using medication terminological systems. J Biomed Inform. 2012;45(4):613-25.

9. McMurtrie D. Medserve [Online]. GitHub; [updated 2018 December 10; cited 2018 December 26]. Available from: https://github.com/AuDigitalHealth/medserve

10. Studer R, Benjamins VR, Fensel D. Knowledge engineering: principles and methods. Data Knowl Eng. 1998;25(1-2):161-97.

11. Haendel MA, Chute CG, Robinson PN. Classification, ontology, and precision medicine. N Engl J Med. 2018;379(15):1452-62.

12. Sohn S, Clark C, Halgrim SR, Murphy SP, Chute CG, Liu H. MedXN: an open source medication extraction and normalization tool for clinical text. J Am Med Inform Assoc. 2014;21(5):858-65.

13. 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.

14. Ferrucci DA. Introduction to “This is Watson”. IBM J Res Dev. 2012;56(3.4):1:1-15.

15. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc. 2010;17(5):507-13.

16. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: A literature review. J Biomed Inform. 2018;77:34-49.

17. Aho AV, Corasick MJ. Efficient string matching: an aid to bibliographic search. Commun ACM. 1975;18(6):333-40.

18. Wang Y, Steinhubl SR, Defilippi C, Ng K, Ebadollahi S, Stewart WF, et al. Prescription extraction from clinical notes: towards automating EMR medication reconciliation. AMIA Jt Summits Transl Sci Proc. 2015;2015:188-93.

19. Li Q, Spooner SA, Kaiser M, Lingren N, Robbins J, Lingren T, et al. An end-to-end hybrid algorithm for automated medication discrepancy detection. BMC Med Inform Decis Mak. 2015;15:37.

20. Uzuner O, 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.

21. Nelson SJ, Zeng K, Kilbourne J, Powell T, Moore R. Normalized names for clinical drugs: RxNorm at 6 years. J Am Med Inform Assoc. 2011;18(4):441-8.

22. Mayo Foundation for Medical Education and Research. MedXN [Online]. GitHub; [updated 2013 December 10; cited 2019 Februrary 23]. Available from: https://github.com/ohnlp/MedXN/

23. Value Set Authority Center. FHIR® terminology service for VSAC resources [Online]. Bethesda, Maryland: National Library of Medicine; [updated 2019 February 6; cited 2019 March 13]. Available from: https://cts.nlm.nih.gov/fhir/

24. Health Level Seven International. HL7 publishes FHIR® Release 4. Ann Arbor, Michigan: Health Level Seven International; [updated 2019 January 2; cited 2019 February 23]. Available from: http://www.hl7.org/documentcenter/public/pressreleases/HL7_PRESS_20190102.pdf

25. Agnew J. HAPI FHIR [Online]. University Health Network; [updated 2019 March 1; cited 2019 March 1]. Available from: http://hapifhir.io/

26. Damerau FJ. A technique for computer detection and correction of spelling errors. Commun ACM. 1964;7(3):171-6.

27. Bahr NJ, Nelson SD, Winnenburg R, Bodenreider O. Eliciting the intension of drug value sets - principles and quality assurance applications. Stud Health Technol Inform. 2017;245:843-7.

28. Uzuner O, Solti I, Cadag E. Extracting medication information from clinical text. J Am Med Inform Assoc. 2010;17(5):514-8.

29. Patrick J, Li M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. J Am Med Inform Assoc. 2010;17(5):524-7.

30. R Core Team. R: a language and environment for statistical computing. 3.5.2 ed. Vienna, Austria: R Foundation for Statistical Computing; 2018.

31. Sohn S, Clark C, Halgrim SR, Murphy SP, Jonnalagadda SR, Wagholikar KB, et al. Analysis of cross-institutional medication description patterns in clinical narratives. Biomed Inform Insights. 2013;6(Suppl 1):7-16.

32. Sittig DF. Grand challenges in medical informatics? J Am Med Inform Assoc. 1994;1(5):412-3.

33. McTaggart S, Nangle C, Caldwell J, Alvarez-Madrazo S, Colhoun H, Bennie M. Use of text-mining methods to improve efficiency in the calculation of drug exposure to support pharmacoepidemiology studies. Int J Epidemiol. 2018;47(2):617-24.

Notices

Disclaimer

This report was prepared by the author in his personal capacity. The opinions expressed in this report are the author's own and do not reflect the view of Integrated Health Information Systems Pte. Ltd. (IHiS).

Conflict of Interest

The author is employed by IHiS while engaged in this research. The author has no other competing interests to declare.

Trademarks and Copyrights

FHIR® is the registered trademark of HL7. The Flame Design mark is the registered trademark of HL7.

Apache UIMA, UIMA, the Apache UIMA logo and the Apache Feather logo are trademarks of The Apache Software Foundation.

MAYO, MAYO CLINIC, and the triple-shield Mayo logo are trademarks and service marks of Mayo Foundation for Medical Education and Research (MFMER). MFMER is the copyright holder of MedXN and MedTagger.

HAPI FHIR is a Java implementation of the HL7 FHIR specification. University Health Network (UHN) is the copyright holder of HAPI FHIR.