Journal of Biomedical Informatics 45 (2012) 626–633

Contents lists available at SciVerse ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Mapping Partners Master Drug Dictionary to RxNorm using an NLP-based approach

Li Zhou [a,b,*], Joseph M. Plasek [a], Lisa M. Mahoney [a], Frank Y. Chang [a], Dana DiMaggio [a], Roberto A. Rocha [a,b]

[a] Clinical Informatics Research & Development, Partners HealthCare System, Inc., Wellesley, United States
[b] Division of General Medicine and Primary Care, Brigham and Women's Hospital Harvard Medical School, Boston, MA, United States

ARTICLE INFO

Article history:
Received 1 August 2011
Accepted 17 November 2011
Available online 28 November 2011

Keywords:
Terminology
Standards
Natural language processing
Medication systems
RxNorm

ABSTRACT

Objective: To develop an automated method based on natural language processing (NLP) to facilitate the creation and maintenance of a mapping between RxNorm and a local medication terminology for interoperability and meaningful use purposes.
Methods: We mapped 5961 terms from Partners Master Drug Dictionary (MDD) and 99 of the top prescribed medications to RxNorm. The mapping was conducted at both term and concept levels using an NLP tool, called MTERMS, followed by a manual review conducted by domain experts who created a gold standard mapping. The gold standard was used to assess the overall mapping between MDD and RxNorm and evaluate the performance of MTERMS.
Results: Overall, 74.7% of MDD terms and 82.8% of the top 99 terms had an exact semantic match to RxNorm. Compared to the gold standard, MTERMS achieved a precision of 99.8% and a recall of 73.9% when mapping all MDD terms, and a precision of 100% and a recall of 72.6% when mapping the top prescribed medications.
Conclusion: The challenges and gaps in mapping MDD to RxNorm are mainly due to unique user or application requirements for representing drug concepts and the different modeling approaches inherent in the two terminologies. An automated approach based on NLP followed by human expert review is an efficient and feasible way for conducting dynamic mapping.

1. Introduction

The adoption of standard medical terminologies is critical to improve semantic interoperability between electronic health records (EHRs). At present, integrating and exchanging medication information is a challenging task since systems often record this information using different terminologies [1,2]. For example, a Computerized Physician Order Entry (CPOE) system might use a local terminology maintained by an institution, a commercial product, or some hybrid of the two [3]. A hospital Pharmacy Information System might use a formulary service terminology [4]. When a medication order gets to the electronic Medication Administration Record (eMAR) and to the dispensing machine [5], medication names may be further modified to a product level description to correspond with the package description at the Food and Drug Administration (FDA)'s National Drug Code (NDC) level [6]. Although effective, unambiguous communication between these systems is critical and the focus of many healthcare safety and quality initiatives [7], it is still true that even within a "closed loop" system, a medication concept may have multiple identifiers and descriptors. An available, reliable standard medication terminology is required to improve the semantic interoperability of information in heterogeneous systems. RxNorm, created and maintained by the National Library of Medicine (NLM), aims to provide a standardized nomenclature that relates itself to terms from commonly used source vocabularies, and to mediate messages between systems not using the same software and vocabulary [8,9]. However, little research has been done to study the mapping between RxNorm and medication terminologies developed at local institutions or other organizations. Such studies are critical for the future adoption and integration of RxNorm in EHRs.

In this study, we present a systematic approach that applies natural language processing (NLP) techniques to facilitate the mapping between RxNorm and an institutional medication terminology. We measure the feasibility and efficiency of our methods via manual review of the mapping results. We also assess the gap between these two terminologies.

2. Background

2.1. Diverse medication vocabularies and RxNorm

Prior to the initial release of RxNorm in 2004, there was an absence of a single, standard, multipurpose terminology for repre-

* Corresponding author at: Clinical Informatics Research & Development, Partners HealthCare System, Inc., 93 Worcester Street, 2nd Floor, Wellesley, MA 02481, United States. Fax: +1 781 416 8912.
E-mail address: Lzhou2@partners.org (L. Zhou).

1532-0464/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jbi.2011.11.006

senting medications [8,10]. In a study conducted in 1996, Kannry et al. [11] mapped 200 pharmacy terms in the Yale Pharmacy Vocabulary to the Columbia Medical Entities Dictionary. The study resulted in a term match rate of 73%. In the late 1990s and early 2000s, multiple collaborative efforts among the HL7 Vocabulary Technical Committee [12], pharmacy system knowledge base vendors [13], the National Library of Medicine (NLM) [14], the Department of Veterans Affairs (VA) [14], and academic institutions [10] lead to the creation and development of RxNorm. Cimino et al. [10] mapped terms from three leading vendors of pharmacy system knowledge bases. The study yielded a 53% match rate. Nelson et al. [14] parsed 70% of the entries in the VA National Drug File (VANDF) to semantic normal forms (SNFs) for clinical drugs. This system of SNFs and their relationships was named RxNorm.

As mentioned above, RxNorm was built by creating normalized drug names based on information from contributing source vocabularies. It not only provides normalized names for clinical drugs, but also links back to the various source vocabularies. After its creation, RxNorm has been increasingly recognized by the biomedical informatics community as an emerging standard for clinical information exchange. Studies were conducted to map local name variants in formularies and pharmacy orders to RxNorm. Peters et al. [15] developed three types of drug-specific normalization rules for mapping local variants in clinical drug names collected from formularies (e.g., state Medicaid formularies) to RxNorm and reached a 45% overall match. The normalization rules include expanding abbreviated word (e.g., "tab" to "tablet"), reformatting specific parts of the drug name (e.g., adding a space between the number and the unit such as transforming "2mg" to "2 mg"), and removing salt modifiers in ingredient names (e.g., "Pseudoephedrine tannate" becomes "Pseudoephedrine"). O'Neill and Bell [16] evaluated the generic RxNorm CUIs (i.e., Semantic Clinical Drug (SCD) and Generic Pack (GPCK)) for representing ambulatory e-prescriptions that were coded using the NDC Directory. Their method for mapping e-prescriptions to RxNorm was mainly based on a NDC-to-CUI mapping derived from RxNorm and a NDC-to-CUI mapping from a medical knowledge base vendor. The STRIDE project [17] developed an algorithm for automated mapping hospital pharmacy orders to RxNorm Ingredient (IN) concept, and achieved a precision of 93%.

The most recent release of RxNorm (July 2011) includes 11 source vocabularies, including FDA NDC and Structured Product Labels (SPLs), VANDF and VANDF_RT, MeSH®, SNOMED CT®, and five vocabularies developed by vendors [8,9]. The Standards and Certification Criteria Final Rule issued by the Department of Health and Human Services (HHS) recognizes any source vocabulary included in RxNorm for representing electronic medication information [18]. However, most of these source vocabularies are oriented to commercial products and information. Many individual institutions and hospitals maintain formularies in a local terminology that meets local user requirements for representing medication concepts. As some comments included in the Final Rule pointed out [18], it is still unclear how well the mapping is between RxNorm and medication dictionaries that are not included in RxNorm or those that are developed at local institutions. It is also unclear what the gaps are and how much effort is required to develop an interface for such a mapping.

2.2. Related work and challenges of medication terminology mapping

Different methods and techniques for mapping terminologies in diverse domains have been widely studied, including lexical methods [17,19,20] semantic methods [21,22], and a combination of both [23]. Mapping medication terminologies presents unique challenges [15]. One challenge is that the structure and drug naming conventions are highly variable between institutions, or for different applications (e.g., CPOE [24] vs. Pharmacy IS [4]). Another challenge is variations in drug names (e.g., ingredient name, generic name, multiple brand names, and longstanding "nicknames" like HCTZ and MS04) and optional drug signature elements (e.g., dose form and strength information), yielding an unlimited number of permutations to represent a term (e.g., "Acetaminophen 325 MG / Oxycodone Hydrochloride 5 MG Oral Tablet [Percocet 5/325]"). The continual evolution of terminologies (e.g., addition, refinement, and obsolescence) also presents a maintenance challenge [25]. The initial mapping and continuous maintenance is costly, if only relying on approaches based on manual review. An automated method would make the process manageable and allow institutional dictionaries to stay "in synch" with the standards.

This paper describes a study that uses an approach based on natural language processing (NLP) to map an institutional drug dictionary to RxNorm and discusses critical challenges and issues related to this mapping. In the following sections, we introduce our local drug dictionary and the NLP system used in this study.

2.3. Partners' Master Drug Dictionary (MDD)

Partners Healthcare is an integrated health care system in the Boston area, founded in 1994 by Brigham and Women's Hospital and Massachusetts General Hospital. It also includes community and specialty hospitals, a physician network, community health centers, home care and other health-related entities. The Partners' Master Drug Dictionary (MDD) is an enterprise-wide drug dictionary, primarily used in inpatient and outpatient CPOE systems. MDD was first developed in 1992. It is maintained by pharmacist knowledge engineers to meet requirements of different academic medical centers and outpatient users. It currently contains 11,697 active drug names. The drug names describe commercially available drugs, as well as investigational drugs and compounded medications that are not commercially manufactured.

MDD medication concepts are composed of three name types: generics, synonyms and misspellings (see Table 1). Generic medication names are typically ingredients or generic descriptors. The generic name is the preferred name in the dictionary for a concept (a medication concept is defined at a "Roll-up" level, as described below). Synonyms may represent a brand name or common name like "Baby Aspirin". The third name type is called a "misspelling". Misspellings are used as pointers to the approved drug names (e.g., "mag sulfate" and "MGS4" point to "magnesium sulfate"). Their use within clinical information systems is confined to making suggestions to improve and correct physicians' order entries at the time of ordering. Misspellings are not displayed or stored in medical records. Each drug name is assigned a unique identifier, called a MedNameID.

A "route group" is associated with the generic name to form the foundation of a MDD medication concept, called a "Roll-up". A medication with an injectable route group may have multiple sub-routes, such as IV bolus, IV infusion or subcutaneous. In addition to order entry, Roll-ups are widely used for clinical decision support (CDS) in Partners EHRs for CDS rules defined at the "routed medication" level (such as drug-food interactions and drug-drug interactions).

Since MDD is originally designed as a CPOE dictionary, it often includes other data elements in the drug names. The drug names may also be de-normalized specifically for the purpose of facilitating prescriber order entry. For example, a specific route (e.g., "nasal"), indication (e.g., "intrathecal use"), dose form (e.g., "oil enema"), or purpose (e.g., "methotrexate for ectopic pregnancy") may be added in the name to allow providers to select the correct drug easily and quickly from a drug name list. It also contains a variety of symbols and spacing variations to improve prescriber recognition.