EVALUATING THE PATIENT-CENTERED AUTOMATED SMS TAGGING ENGINE (PASTE): NATURAL LANGUAGE PROCESSING APPLIED TO PATIENT-GENERATED SMS TEXT MESSAGES

By

Shane P. Stenner

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Biomedical Informatics

August, 2011

Nashville, Tennessee

Approved:

Professor Joshua C. Denny

Professor Kevin B. Johnson

Professor Nancy M. Lorenzi

Professor S. Trent Rosenbloom

Professor Stuart T. Weinberg

ACKNOWLEDGEMENTS

I would like to thank my thesis committee for their guidance throughout this work. I am grateful for my advisor, Josh Denny, who graciously directed and encouraged me through much of this project. I am especially thankful for Kevin Johnson, who served as my advisor early on, helping form the beginning stages of this research and expertly guiding me on the path of my informatics training. Nancy Lorenzi, Trent Rosenbloom, and Stuart Weinberg have shared their expertise as well throughout the course of this work and I am appreciative of their assistance.

I would like to thank Hua Xu for sharing MedEx resources that helped improve the outcome of this work. I would also like to thank the Department of Biomedical Informatics, including the students, faculty, and staff, for their constant support. Thanks particularly to Rischelle Jenkins and Teresa Gillespie who have provided outstanding administrative support and daily smiles. This work was funded by a National Library of Medicine training grant (#T15 007450).

Finally, I am grateful for my family and friends and for their amazing encouragement. My parents, Jack and Lynn Stenner, especially have supported and cheered me on through many years of education and training. I am blessed to have such wonderful, loving family and friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	Definition
ADHD	attention deficit hyperactivity disorder
API	application programming interface
AHRQ	Agency for Healthcare Research and Quality
MDI	metered-dose inhaler
MedLEE	Medical Language Extraction and Encoding System
MMH	MyMediHealth
NDF-RT	National Drug File – Reference Terminology
NLP	natural language processing
PASTE	Patient-centered Automated SMS Tagging Engine
PHR	personal health records
REST	representational state transfer
RWJF	Robert Wood Johnson Foundation
RXCUI	RxNorm Concept Unique Identifier
SCOWL	Spell Checker Oriented Word List
SMS	short message service
VUMC	Vanderbilt University Medical Center
XML	extensible markup language

CHAPTER I

INTRODUCTION

Administration of medical therapies is a key component in the management of nearly every disease. For patients with chronic disease, non-adherence to “controller” medications and the use of acute “rescue” medications for exacerbations are two factors that greatly influence disease control and subsequent medication management. For patients with acute disease, over-the-counter medication use or a visit to an acute clinics that prescribes medications without knowing the patient’s complete history might affect care. Dynamic systems that allow patients to manage medications and inform healthcare providers of medication administration may allow for improved recognition of non-adherence, new medication use, and rescue medication use. Mobile technologies, such as short message service (SMS) text messaging, provide a platform for electronic patient-centered medication management and an opportunity to implement guideline-based support for patients (1-5). There are over 300 million cellular phone subscribers in the United States who send over 2.1 trillion text messages per year; almost every household in the US has at least one cellular phone and over 26% are wireless-only households (2, 6). Ownership and use of cellular phones is as prevalent among those from a lower socioeconomic status as among those from the general population (4, 5); thus, cellular phone technologies may provide an opportunity to significantly decrease healthcare disparities (3, 5).

Pilot studies have demonstrated the feasibility of patient-centered electronic medication management systems with the ability to remind patients about scheduled medications via cell phone SMS text alerts (7-9). In addition to improved guideline-based care, medication management systems for patient use have the potential to intercept drug interactions, stop unintentional medication overdoses, prevent improper scheduling of medications, and to gather real-time data about symptoms, outcomes, and activities of daily living. However, a major challenge to a model for self-care utilizing bi-directional text messaging is the need to process text messages into an accurate computable representation that could be subsequently used by other systems. To date, there has been little published work about this process.

The goal of our project, called the Patient-centered Automated SMS Tagging Engine (PASTE), is to develop a generalizable toolkit for extracting and tagging medication information from patient-generated text messages. We hypothesize that this new natural language processing (NLP) system, PASTE, can accurately extract “medication concepts” and “medication action concepts” from patient-generated messages. System performance will be judged in comparison to medication messages that have been manually tagged by physician reviewers.

CHAPTER II

BACKGROUND

MyMediHealth

MyMediHealth (MMH) is a medication management system created at Vanderbilt University Medical Center (VUMC) that includes a medication scheduler, a medication administration record, and a reminder engine that sends text messages to cell phones (9). A product of Robert Wood Johnson Foundation's (RWJF) Project HealthDesign, MMH was funded through support from RWJF and the Agency for Healthcare Research and Quality (AHRQ). Using MMH, patients can add medications to an online medication list and schedule reminder SMS text messages to be sent to a mobile phone. After a medication reminder message is received, the patient can respond with a two letter command via text message to indicate if they administered the medication ('TO' for took) or did not administer the medication ('SK' for skipped). Figure 1 shows a screenshot of MMH when adding a new medication or scheduling a medication reminder. Figure 2 shows the online medication list and a view of upcoming medication doses in MMH.

Asthma as a Prototypic Chronic Disease for Mobile Technologies

The ongoing AHRQ-funded MMH study is primarily concerned with improving adolescent patient adherence to scheduled daily preventative asthma medication (e.g., inhaled corticosteroids). Patient non-adherence to medical therapies is a major challenge in the management of chronic disease (10-13). In patients with asthma, for example, adherence has been found to be between 30% - 70% (14). Patient self-management in asthma, often assisted by using paper-based "asthma journals" or

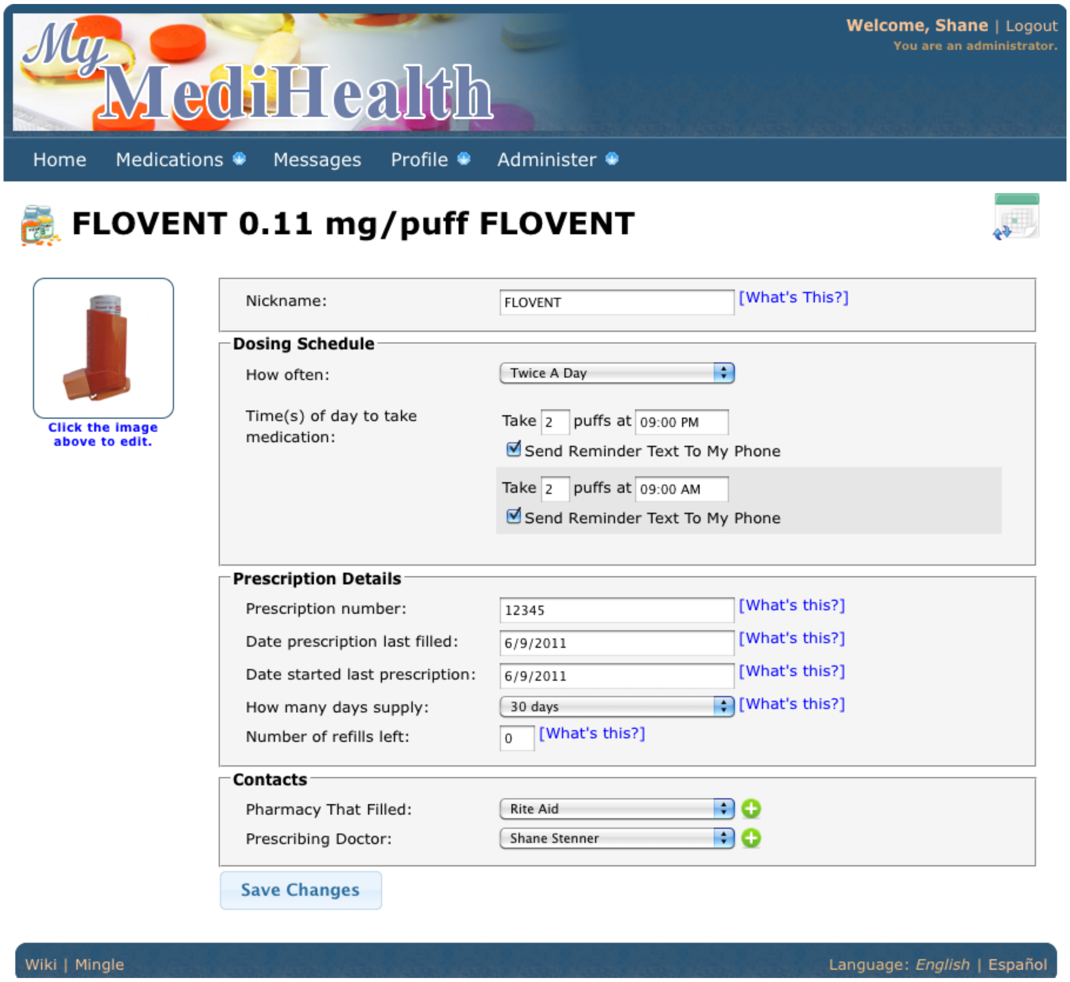

Figure 1. MyMediHealth screenshot showing scheduling of a new medication and reminder

Patients can add medications to their medication list and schedule text message reminders to be sent to their mobile phone.

Figure 2. MyMediHealth screenshot showing medication list and upcoming doses
Patients can view their online list of medications, including images of their medications, and see various calendar views of upcoming medication doses.

self-management plans, is a proven method of improving adherence to guideline-based asthma care – decreasing emergency department visits and increasing medication adherence (15). An electronic system for self-management of disease could capture the frequency and pattern of a patient's "as needed" medication use; these data could be interpreted as a proxy signal for both acute and long-term medication management needs. Beyond asthma, this type of generalizable medication management system could beneficial to patients with diseases where "as needed" medications are more strongly correlated with primary outcomes, such as diabetes mellitus and congestive heart failure. For the purposes of the PASTE and the MMH projects, asthma represents a prototypic chronic disease with well-established clinical guidelines that provide useful examples of acute and chronic disease management. However, the common goal of PASTE and MMH is to provide dynamic, interactive medication management for all patients, regardless of specific disease process.

A recent systematic review of cellular phone use in a variety of healthcare delivery interventions found significant improvements in medication adherence, asthma symptoms, hemoglobin A1c levels in diabetics, stress levels, smoking quit rates, and patient self-efficacy (8, 16-22). Furthermore, cellular phones interventions have lowered missed appointment rates, decreased diagnosis and treatment times, and improved teaching and training of patients (7). Online healthcare applications that allow patients to interact and communicate with their healthcare providers or track their health data, such as patient port sites and personal health records (PHRs), and healthcare applications for mobile devices are increasing in popularity among patients and providers (23-28). These technologies provide new opportunities for increased access to care, more efficient care, and improved clinical outcomes through patient adherence to recommended clinical guidelines (13). In asthma care, for instance, guideline-based self-management plans and medication scheduling decision support could be adapted for delivery through any or all of these modalities. A 2007 systematic literature review of biomedical informatics applications for asthma care by Sanders and Aronsky found that few studies have evaluated the impact of using computerized systems to implement asthma care guidelines (1). While Ostojic showed improvements in asthma symptoms and decreased need for controller medications through the addition of weekly adjustment of treatment using text messaging, not all informatics applications in asthma have demonstrated improved outcomes (8). Porter describes another patient-centered guideline-based technology, the "Asthma Kiosk", which collected data in the ER from parents of children with asthma and was aimed at promoting collaborative decision support with emergency department physicians. The project was successfully deployed and well received by patients but had little clinical impact because of poor provider participation (29, 30). These studies and others have demonstrated that patients and their families make relatively sophisticated medication decisions regarding need for medications and are willing to interact with systems that help them manage their medications in a guideline-based manner (31).

Although MMH is designed for scheduled medications, it requires unprompted text-based communication to record the administration of certain medications, such as those that are taken on an as-needed basis. For example, a patient with asthma who was exposed to an environmental trigger might send a text message to inform the medication management system that he/she used an albuterol inhaler (e.g., 'wheezing took 2 puffs'). This text message is considered "unprompted" because the patient spontaneously sent it to the medication management system and not in response to a reminder text message. Unprompted text-based communication with patients using natural language could engage patients in their healthcare but presents unique NLP challenges. Patient messages cannot be expected to contain structured or complete medication information as one might find in a clinical document (e.g., 'albuterol MDI 2 puffs inhaled'). Additionally, patients may communicate using medication names (brand or generic) or using nicknames for medications (e.g., 'my puffer').

Natural Language Processing

NLP systems identify probable structured concepts contained in narrative text (32). There are many examples of systems that apply NLP techniques to a variety of clinical texts, including MedLEE (Medical Language Extraction and Encoding System), SymText, and the KnowledgeMap concept identifier (33-52). A common approach that NLP systems employ is a process of sentence identification followed by syntactic and/or semantic parsing, followed by lexicon matching (32). Another system, MedEx, was shown to extract medication information with high accuracy from clinical notes using a method of semantic tagging, regular expressions, and rule-based disambiguation components combined with a parser (53). None of these systems were designed to be used with patient-generated text of the informal style contained in typical SMS text messages. NLP challenges unique to text message communication include common use of ad hoc abbreviations, acronyms, phonetic lingoes, improper auto-spell correction, and lack of formal punctuation.

Statistical NLP systems are sensitive to the domain on which they are trained (32). As SMS-based communication with a medication management system is a new domain, existing statistical systems are unlikely to perform well with patient-generated medication messages. Similarly, rule-based NLP systems are not likely to be tuned to the differences in SMS messages compared to traditional biomedical text. Future comparison of PASTE and an existing NLP system could demonstrate these anticipated differences in performance. While models exist for text message normalization, including dictionary substitution and statistical machine translation approaches, we are not aware of any publications that describe an approach specific to patient text messages or to text messages in the domain of medicine (54-59).

CHAPTER III

PILOT STUDY

Introduction

Automating interpretation of text messages containing medication information requires extraction of medication concepts and desired medication actions from patient messages. Examples of important medication actions include: administering a medication, missing or skipping a medication dose, starting or stopping a medication, or canceling a medication reminder. Inclusion of multiple sets of medication/action tuples, contextual ambiguity, and formatting/spelling challenges can each hinder accurate interpretation of SMS messages for automated systems. Patient medication messages can contain multiple sets of medication/action tuples and temporal references. For example, a message could include 'took 2 claratin this am but 4got advair'. This task is further complicated by contextual ambiguity, spelling errors, and phonetic lingoes. Contextual ambiguity refers to instances where information can be understood in more than one way but the context of the information may help resolve the ambiguity. For example, '2' could mean the quantity 2, as it does in the previous example, or it could be an abbreviation of the words 'to' or 'too'. This type of SMS-specific ambiguity is not defined in formal biomedical lexicons, nor is anticipated by current biomedical NLP systems. Temporal information ('this morning' vs. 'now') is important to scheduling and managing medications as well. In the example above, 'am' is used as an abbreviation for 'morning' in reference to taking Claritin®, which the patient also misspelled. Furthermore, 'am' is implied to refer to the time that the patient '4got' (forgot, a phonetic lingo) to take Advair®. Our goal for this pilot evaluation was to develop a system that accurately extracts medication information and administration-related actions from patient SMS messages. This was the first step toward a larger goal of building a system to accurately extract medication information including name, quantity, temporal relationships, and qualitative associations.

System Description

The PASTE webservice uses NLP techniques, custom lexicons, and existing knowledge sources, such as the National Library of Medicine's RxNav webservice, to extract and tag medication concepts from patient text messages. RxNav is a "browser for several drug information sources, including RxNorm, RxTerms and NDF-RT" (60). Using a custom lexicon of "action concepts", PASTE also labels each medication concept with a relevant action corresponding to the existing functions available in MMH:

- *Medication Administered* – the patient took the medication
- *Medication Canceled* – the patient stopped the medication or does not want to be reminded to take the medication any longer
- *Medication Alarm Snoozed* – the patient missed or skipped a medication dose or would like to delay the reminder until later
- *Red Alert* – the patient mentioned a word or phrase that potentially could be dangerous

Once the text message is processed and tagged, structured extensible markup language (XML) is returned. In future versions of MMH, the system will process the XML using a series of rules to record medication administration events, schedule medications, snooze medication alarms, and perform other medication-related tasks. However, for the purpose of this pilot study, the XML messages were simply stored for later review.

The PASTE system and MMH are both hosted on VUMC servers; all scheduling data, knowledge bases, and lexicons are stored in MySQL databases. All message data is collected electronically and is transferred directly from the MMH system to the PASTE system where it is stored in a MySQL database. Both systems were developed and tested to ensure accurate recording of data. Figure 3 shows an overview of the PASTE system and how it interacts with the MMH system.

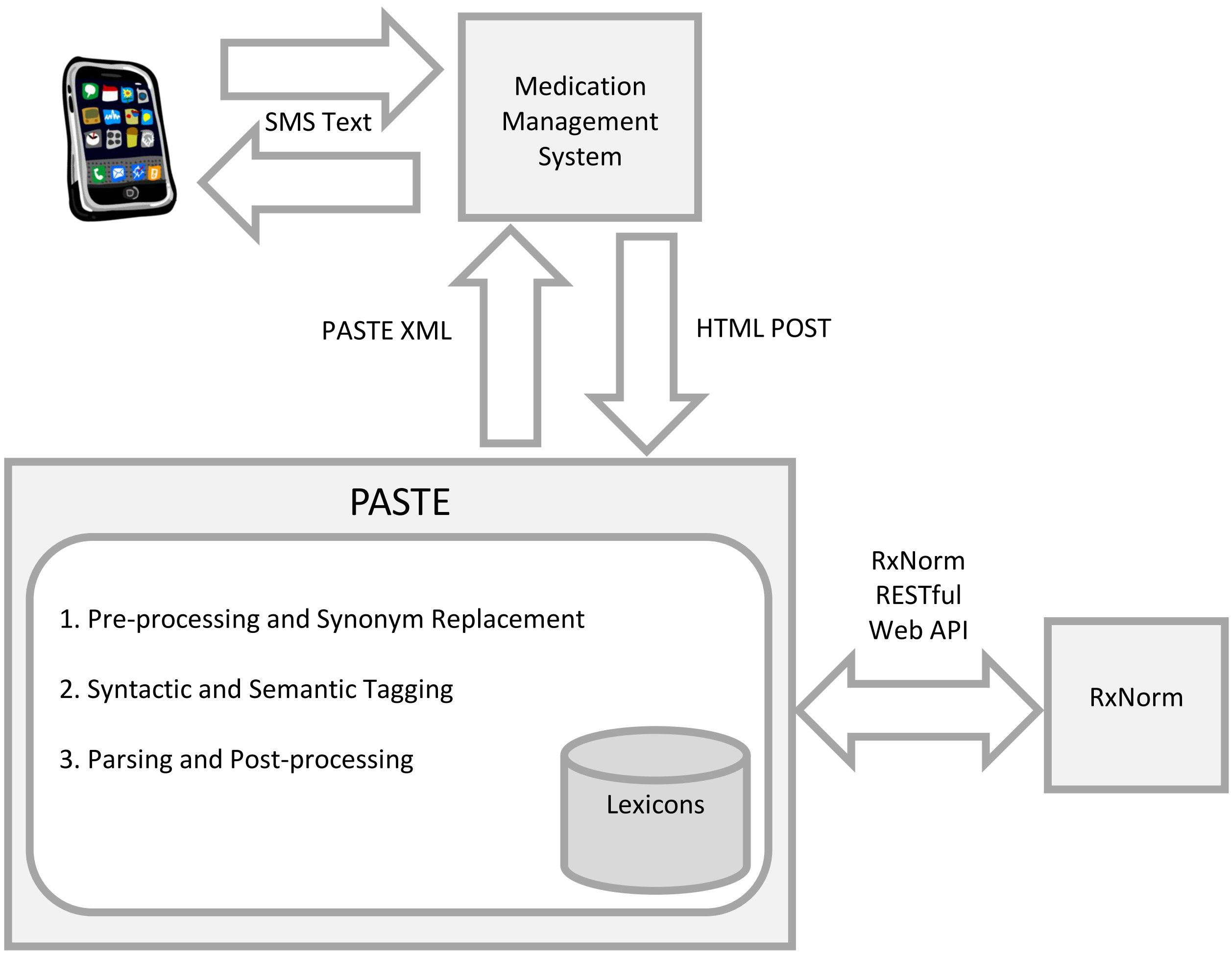

Figure 3: An overview of PASTE and the medication management system.

We define a *medication concept* as a medication name (either generic or brand) that is found in RxNorm, a general medication term (e.g., 'pill', 'med', 'inhaler'), or a candidate medication that is algorithmically matched to a drug concept in RxNorm but is not an exact match (e.g., a misspelled medication name). We define *action concept* as a medication-associated term (e.g., 'took', 'forgot', 'cancel'). Other contextual information such as temporal information was not included in this prototype. RxNav was the only existing system that was utilized as part of PASTE. The prototype system was developed without an extensive test set due to the lack of an easily accessible corpus of patient-generated medication-related text messages.

Pre-processing and Synonym Replacement

The goal of the pre-processing step in PASTE is to remove unnecessary whitespace and symbols in preparation for the lookup tagger. A custom lexicon of over 1,200 common text message abbreviations, acronyms, and phonetic lingoes was constructed from published lookup tables created from CorTxt, a freely available corpus of over 11,000 English text messages (61). This lexicon was combined with other online lists of text message abbreviations and their meanings to form the final lexicon and duplicate entries were removed. A strategy of replacing synonyms was employed to first convert text messages to standard English before tagging. In the case where multiple synonyms were found for a single SMS abbreviation, the most common synonym was manually selected and included in the lexicon or the abbreviation was left unchanged. We plan to address these cases of ambiguity in the future by expanding semantic tagging and contextual strategies, informed by this and future studies. This dictionary substitution of abbreviations, acronyms, and phonetic lingoes is the only normalization performed in the pilot version of PASTE.

Syntactic and Semantic Tagging

Since patient preferences for how to refer to medications are unknown, PASTE was designed to recognize both formal and non-formal references to medications. The general strategy used to tag medications was first to tag known medication names, then tag parts of speech, and finally tag any words that remain untagged as candidate medications. In part because of the variety and poor grammatical structured expected in SMS messages, we did not employ a typical part-of-speech tagger for PASTE. Instead, we assigned parts of speech deterministically using a lexicon and regular expressions. We created the parts of speech lexicon files using freely available online word lists and from an existing medical word list used by KnowledgeMap, another NLP system for biomedical text used at VUMC (36).

A semantic tagging approach was employed for finding medication information, including medication names, general medication terms, and medication related actions. For medication names we chose to use RxNav, a free webservice hosted by the National Library of Medicine that includes drug information from multiple sources, including RxNorm, RxTerms and NDF-RT (62). PASTE sends each non-identified word in a medication message to RxNav via the RxNorm RESTful Web application programming interface (API) and tags each verified medication, including the medication RxNorm Concept Unique Identifier (RXCUI) number in the tag (63). Data sent to RxNav does not contain and cannot be directly linked to any identifying patient information. A lexicon of general medication terms (e.g., ‘pill’, ‘antibiotic’, ‘meds’) was constructed starting with medication forms (e.g., ‘capsule’, ‘lozenge’, ‘suppository’) and then expanded by iteratively adding medication term synonyms from a thesaurus search of existing terms. Similarly, medication classes were included in the list and expanded using a thesaurus search. Semantic tagging of medication action terms used a simple dictionary lookup approach. The medication actions lexicon consists of 140 actions terms that were manually mapped to supported medication management system functions. Finally, all remaining untagged words are labeled candidate medications and the top (most relevant) medication suggestion from the RxNorm Web API function getSpellingSuggestions is included in the tag. Medication suggestions are not evaluated for accuracy after receipt from RxNav.

Regular expressions are used to determine "conceptual segments" by tagging conjunctions (e.g., 'and', 'but', 'nor', 'or', 'yet'). These conceptual units can represent formal or informal independent clauses, which we allow to contain a single medication and action pair. With all tagging complete, PASTE parses messages based on these unit boundaries, which we call "segments." We use regular expressions to find verified medications, actions, and negation signals (e.g., 'didnt tk albuterol') in each segment.

"Conjuctive regularization" of conceptual segments occurs prior to summarization of PASTE's final XML output. Actions and negation are appropriately copied to subsequent segments when medications are linked by a conjunction and the latter medications do not already have linked actions. For example, 'used advair but 4got 2 take zyrtek & my steroid' would be three segments separated by the conjunctions 'but' and '&'. In each segment a medication concept is found (e.g., 'advair', 'zyrtec', 'steroid') and an action term is tagged in the first two segments (e.g., 'used', 'forgot'). In this case, the preceding action term ('forgot') would be copied to the last segment. Punctuation was not used in the pilot version of PASTE to determine unit boundaries.

The final PASTE output from this example is shown in Figure 4 along with intermediate steps.

Methods

The goal of this pilot study is to evaluate the accuracy of PASTE for extracting and tagging medication concepts and action concepts from sample text messages. We gathered a sample corpus of medication messages to be used for training and testing. A group of 47 healthcare professionals and 13 non-healthcare professionals were asked to submit sample medication messages anonymously via a website. We required that all users owned mobile phones and reported regular use of SMS messaging on their phone. They were instructed to include abbreviations, spelling, and punctuation as if they were sending a text message. We instructed them to be creative and to send any kind of medication-related message that they thought relevant. Sixteen volunteers submitted a total of 130 medication messages to PASTE. On the website, users were able to see real-time results of tagging and a sample text message reply from MMH. Participants were not required to be currently taking a medication.

Eighty messages were randomly selected and used as a training set. Improvements were made to the system by manual analysis of the training set. After training, the remaining fifty medication messages were used as the test set for evaluation. An internal medicine physician manually reviewed the fifty test set medication messages and annotated medication names, medication terms, and action terms. The same fifty messages were processed by PASTE and the structured output was manually compared to the expert review gold standard described above. Precision (P), Recall (R), and F-measure (F) were calculated for each type of medication data, where P=TP/(TP+FP), R=TP/(TP+FN), and F=2PR/(P+R), where TP stands for True Positive, FP stands for False Positive, and FN stands for False Negative. Precision is the proportion of cases that PASTE classified as positive that were positive in the gold standard (equivalent to positive predictive value). Recall is the proportion of positive cases in the gold standard that were classified as positive by PASTE (equivalent to sensitivity). F-measure is the harmonic mean of precision and recall (64). As a composite score of precision and recall, F-measure reflects the reliability or accuracy of the system's performance compared to the gold standard.

<u>Sample Input</u>

'used advair but 4got 2 take zyrtek & my steroid'

1. <u>Pre-processing and Synonym Replacement</u>
used advair but ***forgot*** 2 take zyrtek ***and*** my steroid

2. <u>Syntactic and Semantic Tagging: Tag medications that are exact match in RxNorm</u>
used ***<medication><candidate_med>advair</candidate_med><verified>true</verified><rxcui>301543</rxcui></medication>*** but forgot 2 take zyrtek and my steroid

3. <u>Syntactic and Semantic Tagging: Tag parts of speech, actions, and general medication terms</u>
<user_action>used</user_action><action>administered</action>
<medication><candidate_med>advair</candidate_med><verified>true</verified><rxcui>301543</rxcui></medication> ***<conj>but</conj>***
<user_action>forgot</user_action><action>snoozed</action> <num>2</num>
<user_action>take</user_action><action>administered</action> zyrtek ***<conj>and</conj> <pronoun>my</pronoun>***
<med_term>steroid</med_term>

4. <u>Syntactic and Semantic Tagging: Lookup untagged words to match other candidate medications (i.e., misspelled medications)</u>
<user_action>used</user_action><action>administered</action>
<medication><candidate_med>advair</candidate_med><verified>true</verified><rxcui>301543</rxcui></medication> <conj>but</conj>
<user_action>forgot</user_action><action>snoozed</action> <num>2</num>
<user_action>take</user_action><action>administered</action>
<medication><candidate_med>zyrtek</candidate_med><verified>false</verified><suggestion>Zyrtec</suggestion><rxCUI>58930</rxCUI>
</medication> <conj>and</conj> <pronoun>my</pronoun> <med_term>steroid</med_term>

5. <u>Parse and Copy Actions to Appropriate Segments</u>
Segment 0:
<user_action>used</user_action><action>administered</action> <medication><candidate_med>advair</candidate_med>
<verified>true</verified><rxcui>301543</rxcui></medication>
Segment 1:
<conj>but</conj> <user_action>forgot</user_action><action>snoozed</action>
<num>2</num> <user_action>take</user_action><action>administered</action> <medication><candidate_med>zyrtek</candidate_med>
<verified>false</verified><suggestion>Zyrtec</suggestion><rxcui>58930</rxcui></medication>
Segment 2:
<conj>and</conj> <pronoun>my</pronoun> <med_term>steroid</med_term>***<action>snoozed</action>***

6. <u>Output:</u>
<PASTE_XML>
<segment>
<action>administered</action>
<medication>
<candidate_med>advair</candidate_med>
<verified>true</verified>
<rxcui>301543</rxcui>
</medication>
</segment>
<segment>
<action>snoozed</action>
<medication>
<candidate_med>zyrtek</candidate_med>
<verified>false</verified>
<suggestion>Zyrtec</suggestion>
<rxcui>58930</rxcui>
</medication>
</segment>
<segment>
<action>snoozed</action>
<med_term>steroid</med_term>
</segment>
</PASTE_XML>

Figure 4. Sample input, intermediate steps, and output of PASTE.
Changes between steps are shown in ***bold italics***.

Results

In total, there were 31 medication names and 21 medication terms identified by expert review. The results of the pilot test of PASTE, including precision, recall, and F-measure, is reported in Table 1. Verified medication names, medication terms, and action terms reached high F-measures of 91.3%, 94.7%, and 90.4%, respectively. The overall medication name F-measure was 79.8%.

Ten medication messages did not contain an action and three messages did not contain a medication concept (either a candidate medication or a general medication term). Only four messages contained more than one medication concept. There were six unique misspelled medications (“suggested”), fifteen unique properly spelled medications (“verified”), and eleven unique “medication terms”.

Table 1. Results of PASTE on 50 medication messages

Finding Type	Total #	Precision (%)	Recall (%)	F-measure (%)
Medication Name	31	76	84	80
Verified	21	84	100	91
Suggested	10	56	50	53
Med Term	21	100	90	95
Action Term	42	88	93	90
Negation	6	75	100	86
Conjunctions	4	80	100	89

Discussion

In this paper, we present a novel patient medication SMS text message tagger, PASTE, and its pilot evaluation. The PASTE system recognizes explicit medication names and medication terms, and discovers the action words associated with each medication instance, when applicable. Although other studies have demonstrated systems that successfully extract medication information from clinical documents using similar approaches, (65-67) this is the first evaluation of a system to do so from patient text messages. Our evaluation demonstrates that a similar methodology can successfully extract medication information from patient-generated medication messages. This evaluation required us to collect patient medication messages for testing and improvement of the system due to the unique attributes of the corpus. Manual review of the errors generated by PASTE revealed a few repeated causes of errors and some interesting insights about requirements of the system. We believe that we can reach F-measures of greater than 90% for all finding types with further development and testing.

A known limitation of RxNorm is that it contains English words, such as ‘allergy’ or ‘thyroid,’ that patients would not likely use to represent medications. In this evaluation, false positive verified medications were typically due to the presence of non-medical words in RxNorm. Previous systems have addressed this issue by comparing text to a list of general English words like the SCOWL list before searching for them in RxNorm (53, 68). We made similar improvements to PASTE and these are described in Chapter V. In the future, the design goals of MMH would be to allow patients to enter user-specified nicknames (e.g., ‘alb’ or ‘puffer’) for their medications, which could be used instead of either medication names or terms. MMH will replace patient-selected medication nicknames with the associated verifiable medication name before sending the message to PASTE. This feature, once implemented, could increase the precision and recall of the system.

False negatives for actions and general medication terms typically occurred because terms were not included in the respective lexicon. A few false positive actions were due to ambiguity of terms that can sometimes be used to indicate an action. For example, a patient might say ‘now’ or ‘ok’ to indicate that they administered a medication. But they might also say ‘Now is not a good time’ or ‘Is it okay if I take this later?’

The synonym lexicon for text message abbreviations, acronyms, and phonetic lingoes also introduced some errors. In one case, the letter ‘K,’ which was part of a medication name (‘polycitra-k’), was replaced with ‘okay’, which was then tagged a medication administration action. In another example, the word ‘an’ was replaced with ‘and’, which added another segment to the message. False positives in negation were due to inclusion of negating words in the message that were not related to the medication action. The current algorithm does not consider proximity to the medication action, other than being in the same segment. Other causes of errors that we plan to address in future iterations of PASTE include allowing for multiword medication names (e.g., ‘cold meds’, ‘Vitamin B12’, ‘Claritin D’), which most often led to partial matches in the test set as the algorithm in this study assumed all medication names were one word.

A cognitive difference in design and patient expectation of the system was encountered during the evaluation. In the initial development of the system, it was assumed that most text messages from patients would be command-like (i.e., a patient telling the medication management system what he/she did with a medication or what he/she wants to do with a medication). However, early experience with medications messages in this evaluation suggests that patients might desire a more conversational experience with the system. Some messages were formulated in the form of a question, (e.g., ‘I forgot to take my meds last night. Should I take 2 today?’). Other messages indicated a desire for functions that are not currently supported by the action lexicon, such as needing refills, wanting to change medication dosages, and wanting to start a new medication.

A limitation of this study is that messages were not real text messages for patients truly interacting with a dynamic system from which they expected a response, and users could view the output of the system after entering the text message. Thus, users may have been more likely to enter types of entries that they thought the system could understand. Nonetheless, we found the users entered a wide variety of messages – only 2 messages contained the same medication/action pair and similar syntax as another message. Another limitation is a single reviewer validated the output from PASTE. Future improvements to PASTE will expand functionality as described above and focus on improved disambiguation through updated lexicons, broadened semantic tagging, and contextual reasoning.

Conclusion

We developed a novel patient medication text message tagger that extracts medication information as part of a mobile phone based medication management system. In this early evaluation we have shown that PASTE accurately extracts and tags verified medication names, medication terms, and action terms, with over 90% F-measure.

CHAPTER IV

EVALUATION STUDY OF PASTE USING PATIENT-GENERATED SMS TEXT MESSAGES

Introduction

In the prior chapter, we described the initial development and a pilot evaluation of PASTE, which demonstrated the feasibility of extracting medication information from patient-generated medication messages. This chapter explains changes made to the PASTE algorithm to broaden its utility and describes a larger evaluation using patient-generated text messages submitted via mobile phones. The goal of this study was to demonstrate that PASTE can accurately extract medication concepts and desired medication actions from actual patient text messages.

System Description

While they may differ in their methodology and degree of syntactic and semantic parsing, NLP systems commonly employ a strategy of sentence identification followed by parsing and lexicon matching (32). PASTE is a webservice that uses a similar approach but employs a series of custom lexicons, existing knowledge sources, and regular expressions to normalize patient-generated text messages and find medication concepts and actions. Challenges specific to this domain were explored in the pilot study (Chapter III) and informed some of the system changes made as part of this study. PASTE was developed in PHP 5.2.14 and MySQL™ (69, 70).

Methods

We enrolled adolescents (12 – 17 years old) and adults (18 years old and above) for a formal evaluation of PASTE. IRB approval was obtained for the inclusion of both groups. Participants were recruited through the use of the VUMC email listserv, flyers, personal recruitment at local community events, and through the use of websites and social media sites Facebook® and Twitter® (71, 72). All parents/legal guardians and participants completed consent forms and a questionnaire. A sample questionnaire is included in Appendix A. Enrollment criteria included: age 12 and over, use of at least one medication at least three or more times per week, ability to communicate in English, and the possession of a cell phone capable of sending and receiving text messages. There were no exclusion criteria.

Once a participant was enrolled in the study, an account was created for them through the pilot medication management system at VUMC, MMH. The patient then received an SMS invitation to subscribe to the text message routing service used by MMH. Figure 5 shows a schema of text message communication during the study. Participants were asked to imagine an online medication management system that allowed them to enter the names and doses of their medications, receive text message reminders about their medications, and interact with the system by sending text messages. Participants were asked to send 50 text messages about their medications over the course of two months. Suggested message topics included: recording the administration of a medication, canceling a medication reminder, and reporting a missed dose. Example messages were provided and participants were instructed to write and spell as they typically would in a text message. Other instructions were to be creative and to use abbreviations, real medication names, and general terms.

All participant medication messages were saved in a MySQL database without being processed by PASTE. A text message response was sent following delivery of each participant message indicating the number of remaining message required to complete the study. Motivating messages were sent after every 5 – 10 text messages received. During the second month of the study participants received daily reminder text messages to send medication messages. These messages varied by day but were uniform across all participants.

Texts were manually reviewed daily for “red flag” content that might indicate serious or immediate harm could come to the participant or others, in which case the parent/legal guardian and/or proper authorities would have been notified. If participants repeatedly sent the same message(s) they were reminded via text message to send varied and unique messages. Repeat messages, defined as the same message sent more than once in 24 hours, were removed from the dataset. Messages that did not include a “medication concept” were also removed from the dataset. Any reference to a medication was considered a “medication concept”, including medication names, misspelled medication names, and general medication terms such as ‘pill’, ‘inhaler’, or ‘meds’. Once 700 messages were obtained using the above process, the patient medication messages were randomized and split into a training set of 400 messages and test set of the remaining 300 messages. The overall study design is shown in Figure 6. We created a gold standard markup of the test dataset to be used as a comparison to PASTE output after system training was completed. All differences in output between PASTE and the gold standard were later adjudicated by a reviewer. The gold standard creation process, system training, and system evaluation are explained below.

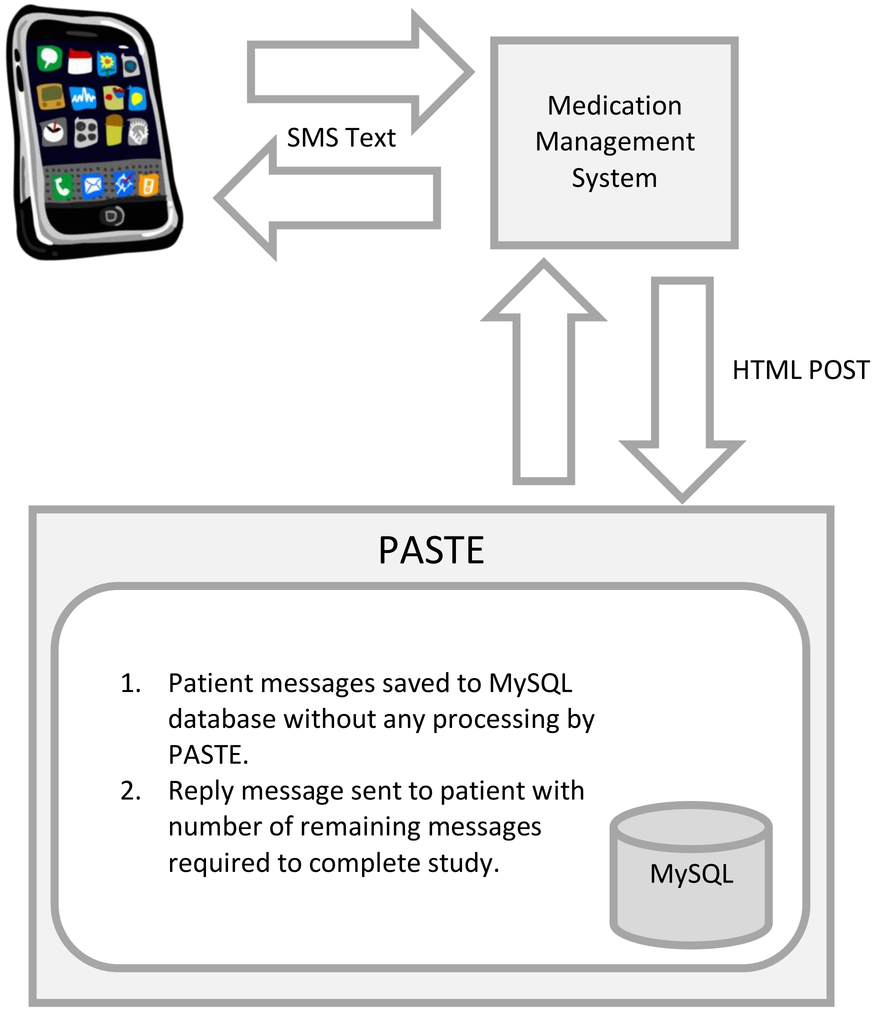

Figure 5. Schema of text communication in PASTE evaluation study. Participants were asked to send 50 messages to PASTE during the study period. They received a text message reply after each message sent as well as periodic encouraging text messages and daily reminder text messages.

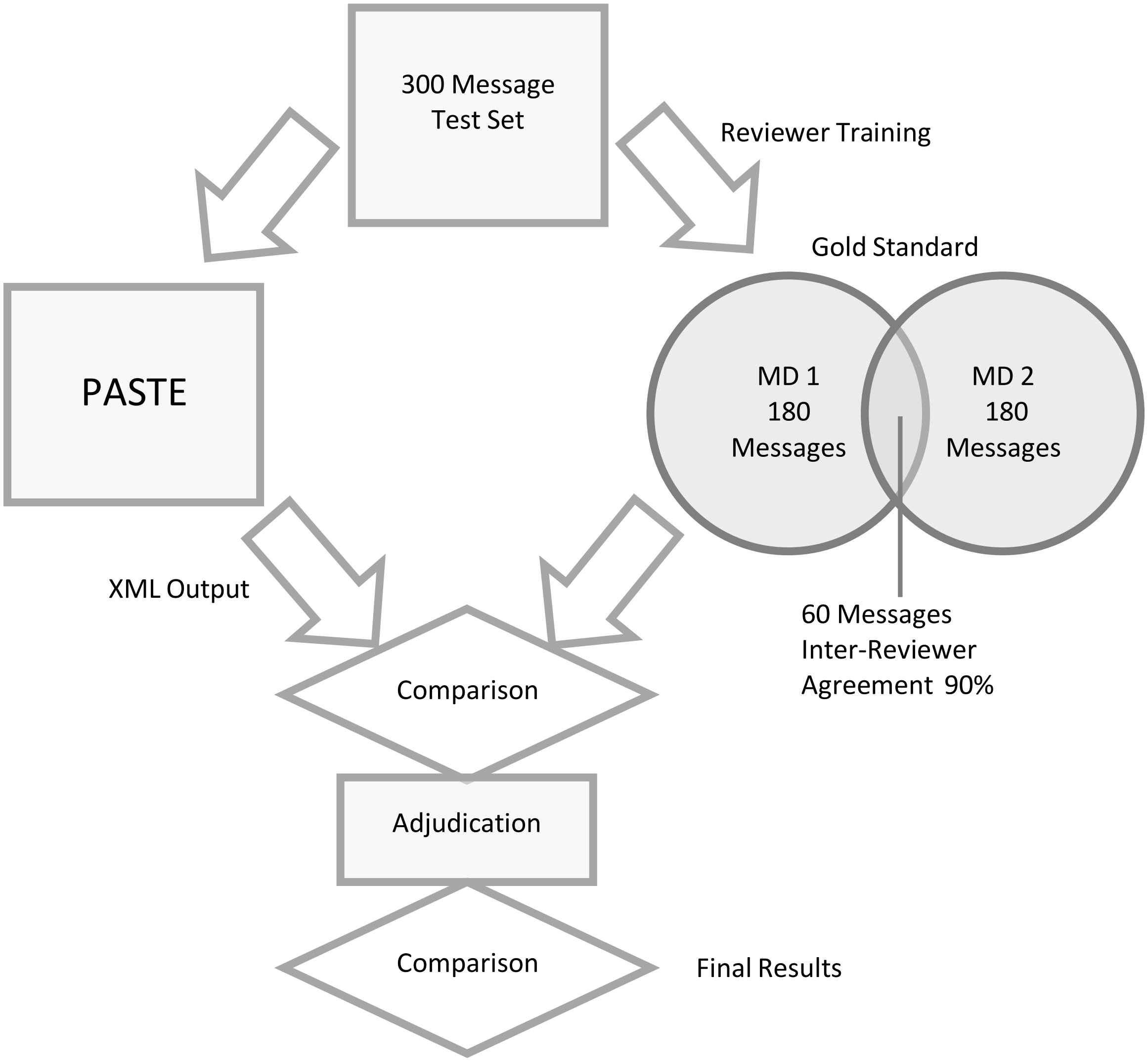

Figure 6. PASTE evaluation study design.

We created a gold standard markup of the test dataset to be used as a comparison to PASTE output after system training was completed. Physician reviewer (MD 1 and MD2 above) were trained until inter-reviewer agreement reached 90%. Then both reviewers marked-up 120 unique messages and 60 common messages, which were compared to validate inter-reviewer agreement in the gold standard. All differences in output between PASTE and the gold standard dataset were later adjudicated by one of the reviewers.

Gold Standard Creation

To create the gold standard test dataset, two physicians independently reviewed and tagged messages for medication concepts and their actions, and classified each medication segment as whether it was a hypothetical statement (defined as a message that includes a medication action but for which the action should not be recorded as completed) or the user was stating a question ('Will I ever be able to stop taking my Adderall for ADHD?'). A training period began with twenty-five messages from the training set that were displayed on a custom web-based interface, shown in Figure 7 below.

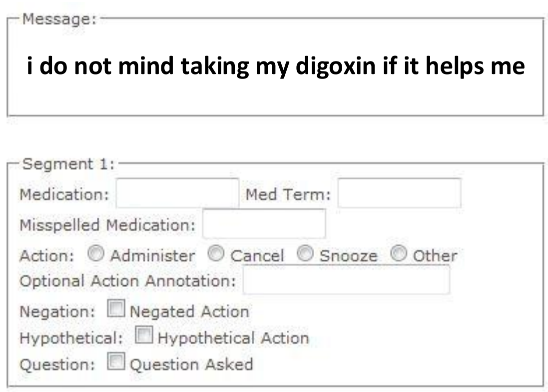

Figure 7. Reviewer interface for gold standard creation.

The physician reviewers were asked to identify the following concepts in each message:

1. Medications: Correctly spelled medication names including brand and generic names. Included anything that the physician thought might be included in RxNorm.
2. Misspelled medications: Medication names that were abbreviated or misspelled
3. Medication terms: General terms that referred to a medication but were not proper medication names or misspelled medication names or which the physician did not believe

would be in RxNorm. Examples include ‘pills’, ‘medications’, and ‘oatmeal bath’.

4. Medication Action: What the patient did with the medication. Options included “Administer”, “Snooze”, “Cancel”, and “Other”. These terms are further defined below.
5. Negation: The presence of any negating term like ‘not’, ‘didn’t’, or ‘never’ referring to a medication or a medication action.
6. Hypothetical: Statements that include “medication action” concepts that should not be recorded as an event. An example would be, ‘I wish I had to take fewer pills’
7. Question: Any phrase that indicates a question is being asked about a medication or a medication action.

The medication action concepts described above were chosen to map to existing supported features in MyMediHealth. “Administer” concepts included any term or phrase to indicate a patient took a medication. An action was labeled “snooze” if a patient missed, forgot, or skipped a medication dose. Actions were labeled “cancel” if a patient was stopping a medication or did not want to be reminded to take a medication any longer. Finally, reviewers could select “other action” when patients expressed actions not among these three options. Messages were subdivided into conceptual “segments” that were bounded by punctuation, conjunctions, or other indications of medication-action tuples.

Following work with the initial training set, the author and two reviewers categorized message tagging results and the latter two elements in the list above, hypothetical and question statements, were added to the review interface. Next, the physician reviewer pair independently reviewed and tagged an additional 25 training set messages. Once inter-reviewer agreement was determined to be sufficiently high (>90%), the 300 test dataset messages were divided among the reviewers, with each receiving 180 messages to independently review and tag. Sixty of the 300 messages (20%) were reviewed by both reviewers to ensure inter-reviewer agreement in the test dataset. The review and tagging tool (shown in Figure 7) automatically generated the gold standard XML for comparison to the output of PASTE.

System Training

Independent of the test set development, the author manually tagged the four hundred training set messages using the same review and tagging tool as the physician reviewers. Training of PASTE was completed using an iterative process of system changes and comparison of system output to the manually tagged training set output. Training continued until the F-measure for training set data exceeded 0.90 and recall and precision could no longer be improved.

Evaluation

PASTE output from the test set was compared to the manually-tagged gold standard XML. We calculated Precision (P), Recall (R), and F-measure (F) for each type of medication data, where P=TP/(TP+FP), R=TP/(TP+FN), and F=2PR/(P+R), where TP stands for True Positive, FP stands for False Positive, and FN stands for False Negative. Precision is the proportion of cases that PASTE classified as positive that were positive in the gold standard (equivalent to positive predictive value). Recall is the proportion of positive cases in the gold standard that were classified as positive by PASTE (equivalent to sensitivity). F-measure is the harmonic mean of precision and recall (64). As a composite score of precision and recall, F-measure reflects the reliability or accuracy of the system's performance compared to the gold standard.

We found that differences between reviewers and between the gold standard and PASTE outputs were often due to human failures to identify certain items (e.g., forgetting to label a medication or general medication term, missing a negation, question, or hypothetical statement) or non-standardized interpretations of categorizations (e.g., labeling implied actions as “administered”, tagging ‘did not take’ as “snooze” or “cancel” instead of “negated” + “administered”). Thus, all differences between the gold standard and PASTE were adjudicated by one of the physician reviewers. The XML display order was randomized for each message and the reviewer was blinded to the source of each output. Differences between messages were highlighted. The adjudication interface (shown in Figure 8) allowed the reviewer to select one of the two XML message markups as correct or enter a completely new gold standard XML markup using the same interface as in the original tagging process.

The adjudication process ensured standardization of message tagging including tagging of medications without explicitly stated actions and tagging of hypothetical actions. For example, a message that says ‘Advair at 5pm’ could imply an administration event (that the patient used Advair® at 5pm). A design decision was made in PASTE to only record explicitly stated actions and so the gold standard XML for the above statement would include a tag for the medication, Advair®, but no tag for an associated action, as one cannot know for certain if the above statement indicates if it was taken, planned to be taken, or missed at 5pm. Additionally, the tagging of axiomatic messages such as ‘I take Claritin for my allergies’ or truly hypothetical messages like ‘I might want to stop my antidepressant’ was standardized as a hypothetical action (as opposed to an “other” action without a “hypothetical” tag). The final gold standard output was then compared to PASTE output to obtain the final adjudicated results. Data were analyzed using Microsoft Excel and Stata 11 statistical software.

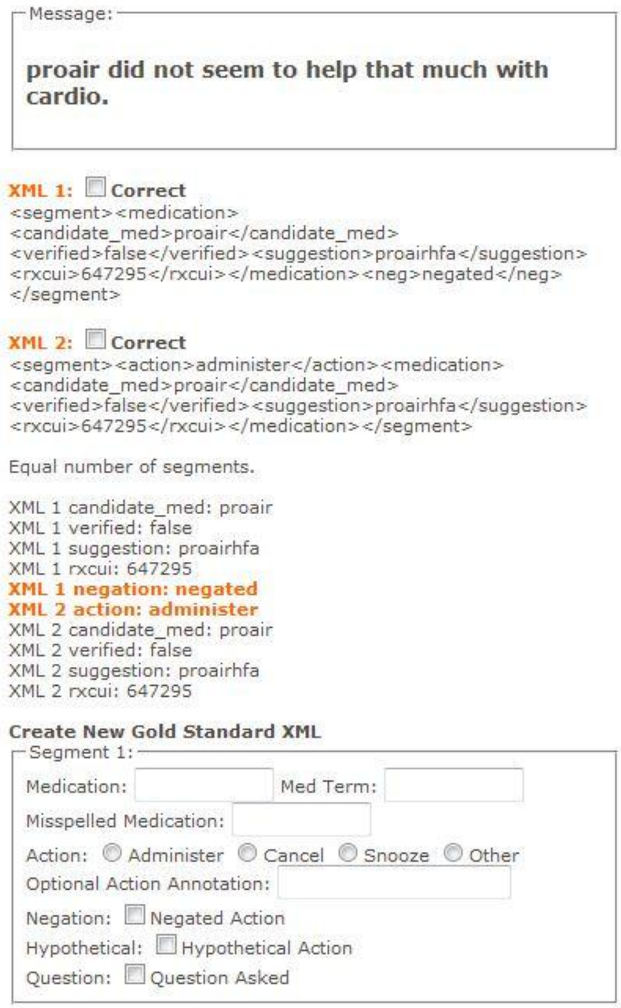

Figure 8. Reviewer adjudication interface.

Differences between the gold standard and PASTE were adjudicated by one of the physician reviewers. XML order was randomized for each message and the reviewer was blinded to the source of each output. Differences were highlighted and the reviewer could select the correct output or create a new gold standard for the message.

Improvements in PASTE System Design

The challenges uncovered during the pilot study of PASTE informed PASTE improvements that were completed before and during the training period of this study. In the pilot study, false positive medication matches occurred due to the presence of non-medication standard English words in RxNorm. Another limitation was the lack of a robust standard English lexicon in the pilot system, leading to false tagging of properly spelled non-medication words as candidate medications. Another NLP system, MedEx, had overcome the false positive RxNorm match problem by creating an exclusion list of terms that should not be sent to RxNorm as possible medications (53). Using the MedEx exclusion list, we were able to address this false positive problem in a similar way. Additionally, we added a spellchecker as a final step before sending any terms to RxNorm for algorithmic matching of candidate medications, which we hoped would decrease false positive medication tags.

We employed iterative changes to PASTE to improve system performance as compared to the manually tagged training set of 400 patient medication text messages. Precision and recall were assessed after each successive change until both measures were greater than or equal to 0.90 and a saturation point was achieved, after which new changes did not result in performance gains.

Normalization of non-standard English in text messages was a priority for decreasing false positive identification of medication candidates. In addition to updating synonym lexicons we created new versions of existing lexicons that had all vowels removed. These new lexicons mapped to the existing lexicons in an attempt to match some ad hoc abbreviations. Regular expressions were also updated to normalize long repeating letters and punctuation. In a real example from the test dataset, 'i take myyyy digoxinnnn everyyy dayyyy(:' was normalized to 'i take my digoxin every day(:'. After normalization, the parts of the medication text message could then be matched to lexicons or RxNorm and tagged appropriately.

The pilot version of PASTE only considered single word medication names and medication terms. We added the ability in PASTE to match multiword terms across all lexicons and handle nuanced medication matching in RxNorm, such as trying both a space and a dash between words in a potential medication name. Additionally, a lexicon was created to replace misspelled medication terms that

RxNorm did not match correctly using the getSpellingSuggestions function of the RxNorm RESTful Web API.

The pilot version of PASTE defined medication segments using only conjunctions. During the training period of this study we added support for punctuation such as commas, periods, and parentheses as bounds for concept segments and improved the functionality of conjunctive regularization, which was described in Chapter III.

During reviewer training and early gold standard creation we recognized the need to identify questions and “hypothetical” or non-action phrases in the gold standard. We also updated PASTE to similarly identify questions and “hypothetical” actions. As an example, the actual patient message shown in Figure 7, ‘i do not mind taking my digoxin if it helps me’, would be tagged as hypothetical since the phrase ‘i do not mind’ indicates that the medication action (‘taking’) is not active. The system should tag the medication action concept as “administer” and “hypothetical”, since the message refers to the action of taking a medication. However, it would be an error to interpret the message as an administration event. Similarly, if a patient asks a question such as, ‘Will I ever be able to stop taking my Adderall for ADHD? I really hope so…’ (an actual example), the medication management system should not record that the patient ‘stop(ped) taking’ Adderall®. We identified questions using regular expressions to find punctuation, as well as question words and phrases, since many questions did not actually end in question marks. Phrases in the training set that identified “hypothetical” medication statements were also matched using regular expressions. We considered a phrase “hypothetical” if it modified a medication action to indicate the action had not been completed, would be completed in the future, or that the action was being referred to but not completed. Although not truly “hypothetical”, this definition included axiomatic statements such as ‘i take my 6mp for my crohns disease’ or ‘i take cetrizine every day.’ Note that in each of these examples, the correct action modality is “administered.”

Results

Out of 19 participants, more were female than male (14 female, 5 male) and more were teens than adults (14 teens, 5 adults). Teens reported sending more text messages per month (1000 – 1499 per month) on average than adult participants (100 – 499 per month). Adults reported taking a higher number of medications on a regular basis (3 or more times per week) than teens.

Table 2. Participant demographics comparing adults and teens

	Overall		Adults		Teens	
Total Participants	19		5	(26%)	14	(74%)
Female	14	(73%)	4	(80%)	10	(71%)
Male	5	(26%)	1	(20%)	4	(29%)
Mean Age (min – max)	20.6	(13 – 63)	36.4	(26 – 63)	15	(13 – 17)
Median Age (years)	16		27		15	
Mean Medications	2.3		2.6		2.1	
Median Medications	2		3		1	
Text messages/month						
Less than 100	2	(11%)	1	(20%)	1	(7%)
100 – 499	9	(47%)	4	(80%)	5	(36%)
500 – 999	–		–		–	
1000 – 1499	1	(5%)	–		1	(7%)
1500 – 1999	2	(11%)	–		2	(14%)
2000 or more	5	(26%)	–		5	(36%)
Family income						
\$20,000 – \$40,000	3	(16%)	–		–	
\$40,000 – \$70,000	5	(26%)	–		–	
More than \$70,000	10	(53%)	–		–	
Declined to answer	1	(5%)	–		–	
Highest education level of participant or parent						
High School	2	(11%)	–		–	
Some College	2	(11%)	–		–	
College Degree	11	(58%)	–		–	
Graduate Degree	4	(20%)	–		–	

Overall, the participants or their parents/guardians were well educated and represented a high socioeconomic group, reflecting the high number of participants who were recruited on the Vanderbilt medical campus – 80% percent of participants or their parents/guardians had at least a college degree and 53% of all participants or their parent/guardian reported and annual household income of more than $70,000 per year. Eight-six percent of teen participants lived in two-parent homes. One teen was home-schooled; all remaining teens attended public schools. Eighty-nine percent of all participants were white. A variety of mobile phone text entry form factors were represented within the study group, including three number keypad phones (16%), seven with full QWERTY keyboards (37%), five with both QWERTY keyboards and touchscreens (26%), and four with touchscreen-only phones (21%). All five adult participants and eight of the 14 teen participants (57%) sent the 50 text messages required to complete the study. Overall, there was a 68% completion rate for study participants. Table 2 shows the demographics of the study participants and compares the adult and teen groups.

Table 3. Comparison of test dataset and overall dataset

	Total	Adults		Teens	
Overall Dataset					
Number Text Messages	700	283	(40.4%)	417	(59.6%)
Range	7 – 77	38 – 77		7 – 54	
Mean	36.8	54.2		32.6	
Median	44	50		34.5	
Test Dataset					
Number Text Messages	300	124	(41.3%)	176	(58.7%)
Range	0 – 41	16 – 41		0 – 22	
Mean	15.6	23.8		12.7	
Median	16	21		15	

Table 3 shows that randomization of the dataset was effective; adults and teens contributed 40.4% and 59.6% of the total number of text messages to the overall dataset respectively compared to 41.3% and 58.7% of the test set. There were 49 unique RXCUIs and 50 unique medication terms in the test dataset. Adults contributed more messages per participant than teens. Table 4 shows that the number of concepts per message and the preferred concept types varied between teens and adults. Overall there were 3.54 tags per message in the gold standard test set (4.22 per message for adults and 3.07 per message for teens). Adults were nearly four times more likely than teens to use a medication name. Teens were seven times more likely than adults to use general medication terms, and three times more likely to send hypothetical statements about their medications.

Table 4. Number of concept tags per message in the gold standard test dataset

	Overall	Adults	Teens
Total Number of Tags	1063	523	540
Number of Messages	300	124	176
Per Message:			
Total Tags	3.54	4.22	3.07
Action	0.78	0.70	0.84
Candidate Medication	0.54	0.94	0.25
Verified Medication	0.54	0.94	0.25
Suggested Medication	0.18	0.26	0.13
RXCUI	0.54	0.94	0.25
Medication Term	0.56	0.12	0.86
Negation	0.08	0.07	0.09
Hypothetical	0.21	0.10	0.30
Question	0.12	0.15	0.10

Table 5 reports the non-adjudicated and adjudicated results overall and by specific concept tag. Precision and recall were improved with adjudication across all concepts except "candidate medication", "verified medication", and "rxcui" tags, where recall was unchanged. The overall F-measure increased from 0.82 to 0.88 with adjudication. F-measures for negation, hypothetical statements, and questions were lower (0.75, 0.72, and 0.76 respectively) despite adjudication. There were fourteen instances of partial matches due to multiword general medication phrases in the gold standard, which contributed to the large gains in precision and recall for "medication term" concepts following adjudication (pre-adjudication precision 0.83 and recall 0.80 increased to 0.96 and 0.84 respectively with adjudication).

Table 5. Comparison of non-adjudicated and adjudicated results by tag

	Non-Adjudicated		Adjudicated	
	Score	**(95% C.I.)**	**Score**	**(95% C.I.)**
Overall				
Precision	0.74	(0.72 – 0.77)	0.85	(0.83 – 0.87)
Recall	0.90	(0.88 – 0.92)	0.92	(0.90 – 0.95)
F-measure	0.82		0.88	
Action				
Precision	0.82	(0.77 – 0.87)	0.89	(0.84 – 0.92)
Recall	0.94	(0.90 – 0.97)	0.96	(0.93 – 0.98)
F-measure	0.88		0.92	
Candidate Medication				
Precision	0.78	(0.71 – 0.84)	0.89	(0.83 – 0.93)
Recall	0.96	(0.92 – 0.99)	0.96	(0.92 – 0.99)
F-measure	0.86		0.92	
Verified Medication				
Precision	0.78	(0.71 – 0.84)	0.89	(0.83 – 0.93)
Recall	0.96	(0.92 – 0.99)	0.96	(0.92 – 0.99)
F-measure	0.86		0.92	
Suggested Medication				
Precision	0.55	(0.43 – 0.67)	0.65	(0.53 – 0.76)
Recall	0.90	(0.77 – 0.97)	0.92	(0.80 – 0.98)
F-measure	0.68		0.76	
RXCUI				
Precision	0.77	(0.70 – 0.83)	0.86	(0.80 – 0.91)
Recall	0.96	(0.92 – 0.99)	0.96	(0.92 – 0.99)
F-measure	0.85		0.91	
Medication Term				
Precision	0.83	(0.75 – 0.88)	0.96	(0.91 – 0.98)
Recall	0.80	(0.73 – 0.86)	0.84	(0.78 – 0.89)
F-measure	0.82		0.90	
Negation				
Precision	0.44	(0.27 – 0.62)	0.65	(0.46 – 0.80)
Recall	0.83	(0.56 – 0.96)	0.88	(0.69 – 0.97)
F-measure	0.58		0.75	
Hypothetical				
Precision	0.51	(0.39 – 0.63)	0.68	(0.56 – 0.79)
Recall	0.62	(0.48 – 0.74)	0.77	(0.64 – 0.86)
F-measure	0.56		0.72	
Question				
Precision	0.46	(0.31 – 0.63)	0.71	(0.55 – 0.84)
Recall	0.79	(0.58 – 0.93)	0.83	(0.66 – 0.93)
F-measure	0.58		0.76	

Table 6. Comparison of PASTE performance on adjudicated adult and teen messages

	Adults		Teens	
	Score	**(95% C.I.)**	**Score**	**(95% C.I.)**
Overall Dataset				
Precision	0.94	(0.92 – 0.96)	0.77	(0.74 – 0.80)
Recall	0.95	(0.93 – 0.97)	0.90	(0.87 – 0.92)
F-measure	0.95		0.83	
Action				
Precision	0.92	(0.85 – 0.97)	0.87	(0.81 – 0.92)
Recall	0.98	(0.92 – 1.00)	0.95	(0.90 – 0.98)
F-measure	0.95		0.91	
Candidate Medication				
Precision	0.99	(0.95 – 1.00)	0.69	(0.58 – 0.80)
Recall	0.97	(0.91 – 0.99)	0.95	(0.85 – 0.99)
F-measure	0.98		0.80	
Verified Medication				
Precision	0.99	(0.95 – 1.00)	0.67	(0.54 – 0.79)
Recall	0.97	(0.91 – 0.99)	0.95	(0.84 – 0.99)
F-measure	0.98		0.72	
Suggested Medication				
Precision	0.87	(0.69 – 0.96)	0.38	(0.32 – 0.65)
Recall	0.93	(0.77 – 1.00)	1.00	(0.82 – 1.00)*
F-measure	0.90		0.55	
RXCUI				
Precision	0.96	(0.91 – 0.99)	0.67	(0.54 – 0.79)
Recall	0.96	(0.91 – 0.99)	0.95	(0.84 – 0.99)
F-measure	0.96		0.79	
Medication Term				
Precision	1.00	(0.77 – 1.00)*	0.95	(0.90 – 0.98)
Recall	0.93	(0.68 – 1.00)	0.83	(0.76 – 0.89)
F-measure	0.97		0.89	
Negation				
Precision	0.67	(0.35 – 0.90)	0.64	(0.41 – 0.83)
Recall	0.89	(0.52 – 1.00)	0.88	(0.62 – 0.98)
F-measure	0.76		0.74	
Hypothetical				
Precision	0.54	(0.25 – 0.81)	0.71	(0.58 – 0.82)
Recall	0.58	(0.28 – 0.84)	0.81	(0.67 – 0.90)
F-measure	0.56		0.76	
Question				
Precision	0.88	(0.64 – 0.98)	0.58	(0.37 – 0.78)
Recall	0.83	(0.59 – 0.96)	0.82	(0.57 – 0.96)
F-measure	0.86		0.68	

*one-sided 97.5% confidence interval

Table 6 reports the adjudicated results of PASTE performance by concept category for both adults and teens. Overall performance was much better for adults compared to teens (F-measures 0.95 and 0.83 respectively). Precision for candidate medications in teen messages was particularly low due to a higher rate of non-standard English by teens, which lead to false positive candidate medications. There were a small total number of "suggestion" tags (55), "negation" tags (25), "hypothetical" tags (64), and "question" tags (36) in the gold standard dataset. Precision and recall were lower for these concepts than then remaining five concepts and the 95% confidence intervals were wide, due to their low frequency (Table 5 and Table 6). Overall performance of PASTE for finding "medication concepts" (a composite of candidate medication concepts and general medication term concepts) and "action concepts" was very good with F-measures of 0.91 and 0.92 respectively (Table 7). System performance in these two main categories was excellent for adult medication messages (F-measures of 0.98 and 0.95 respectively) and good for teen medication messages (F-measure 0.85 and 0.91 respectively).

Table 7. Comparison of composite medication concepts and action scores between groups

	Overall		**Adults**		**Teens**	
	Score	**(95% C.I.)**	**Score**	**(95% C.I.)**	**Score**	**(95% C.I.)**
Medication Concepts						
Precision	0.92	(0.88 – 0.95)	0.99	(0.96 – 1.00)	0.85	(0.79 – 0.90)
Recall	0.90	(0.86 – 0.93)	0.96	(0.91 – 0.99)	0.84	(0.78 – 0.89)
F-measure	0.91		0.98		0.85	
Actions						
Precision	0.89	(0.84 – 0.93)	0.92	(0.85 – 0.97)	0.87	(0.81 – 0.92)
Recall	0.96	(0.93 – 0.98)	0.98	(0.92 – 1.00)	0.95	(0.90 – 0.98)
F-measure	0.92		0.95		0.91	

Discussion

In this paper, we describe an expanded evaluation of a novel patient medication SMS text tagger, PASTE. In the pilot evaluation, we previously demonstrated the feasibility of using a combination of semantic tagging and regular expression-based approaches to successfully extract medication information from patient-generated medication messages. Patient-generated medication messages via SMS text are a new domain, which required the collection of actual text messages from patients for further improvement and testing of the system. Overall performance was improved from the pilot version of PASTE, with F-measures for “medication concepts” and “action concepts” remaining over 90%, using a much larger set of messages entered directly from patient phones from both teens and adults. Review of successes and failures reiterate some of the same challenges uncovered in the pilot evaluation. In addition, new categories of input were found in this evaluation that informs development of future SMS text message systems for medication management.

Adjudication resulted improved performance measures due to standardization of concept tagging between 2 reviewers, correction of human error in the gold standard, and correction of partial matches. For example in the message ‘allergy meds makes me drowsy. Sleeeeepy!’ PASTE matched ‘meds’ as the general medication term but the gold standard contained ‘allergy meds’ as the general medication term. In another example, PASTE corrected ‘cntrl’ in ‘I jst took mi birth cntrl’ to ‘control’ and then matched ‘birth control’ as the general medication term. It was not counted as a correct match in the non-adjudicated results because the gold standard contained ‘birth cntrl’ as the general medication term. This example also demonstrates a success of normalization using synonym replacement of terms without vowels. Another success of normalization, but with regular expressions instead of lexicons, is demonstrated by the example ‘i take myyyy digoxinnnn everyyy dayyyy(:’ PASTE normalized this message to ‘i take my digoxin every day(:’ and matched digoxin in RxNorm as a candidate medication concept. This match was also discounted in the non-adjudicated results as the gold standard included ‘digoxinnnn’ as the medication concept.

As in the pilot study, false positive candidate medications were a major challenge for the system. There are two main reasons why this challenge remains. First, the algorithm for PASTE matches and tags as many terms as possible and then considers all terms left untagged as potential misspelled medication candidates. Therefore, any non-standard English term that is not matched in any of PASTE's lexicons is considered a candidate medication term. For example, in the patient message 'took my pills this morning and now i am eating B-fast w/a cute RUSSIAN waiter! i<3NewYork YAY!' PASTE incorrectly tagged 'B-fast' as a candidate medication, which was sent to RxNorm and matched to the medication concept 'Pro-Fast' (RXCUI 705879). The second reason that false positive medication concepts remain a challenge is that actual misspelled medications terms or abbreviations are not well matched by the existing RxNorm getSpellingSuggestions web API function. Even when PASTE correctly identified a misspelled medication, the suggested medication name from RxNorm was frequently incorrect. A few examples of incorrect medication suggestions include:

PASTE misspelled medication: 'ibprfn' (ibuprofen)
RxNorm medication suggestion: 'ibren RXCUI 352708'

PASTE misspelled medication: 'oscal' (Os-Cal®)
RxNorm medication suggestion: 'docal RXCUI 216689'

PASTE misspelled medication: 'prvcd' (Prevacid®)
RxNorm medication suggestion: 'uracd RXCUI 539389'

PASTE misspelled medication: 'tenix' (Tenex®)
RxNorm medication suggestion: 'atenix RXCUI 151387'

PASTE misspelled medication: 'zerrontin' (Zarontin®)
RxNorm medication suggestion: 'neurontin RXCUI 196498'

Furthermore, the exclusion list approach of removing non-medication terms before sending candidate medications to RxNorm was successful at decreasing false positives, but failed on at least one occasion when a non-medication RxNorm term was included in a message but was not on the exclusion list. In the patient message 'I like the color of my zerrontin. it's a pretty orange.<Broken~|~Mirror>' the term ‘orange’ was matched in RxNorm as a medication with the RXCUI 892565.

The three previous patient message examples highlight a new challenge that was encountered in this study: text message signatures and emoticons. These were typically handled by regular expressions but future versions of PASTE could include tagging of emoticons to be used as qualitative associations of medication messages. Another new challenge encountered during this study was false negative matches due to auto correct or mistyping. In one patient message ‘I took my need today’ it is likely that the patient intended to type ‘I took my *med* today’ (and considered as such in the gold standard). When phones autocorrect text to standard English terms it can lead to false negative matches. This currently does not appear to be a major problem but future incorporation of a more sophisticated part of speech tagger and contextual reasoning into PASTE combined with a reverse autocorrect lookup table for medication names and terms could be one approach to overcoming the challenge.

As might be expected, PASTE performed best when patients used standard English terms and properly spelled medication names. As an example, the patient message ‘Just took mornin meds (oscal, prednisone, plaquinil, and folic acid)’ was tagged perfectly by PASTE. Although ‘mornin meds’ was not in the general medication term lexicon, the term was ignored because exact match RxNorm medications were found. PASTE also correctly regularized the “administered” action concept to each of the medication terms listed in the parenthetical phrase. PASTE’s superior performance on messages with formal medication names and properly spelled medication names explains, in part, the better performance for messages from adults, who preferred to use formal medication terms. Ad hoc abbreviations, misspelled medications, and non-standard English were more common in teen text messages and remain a major challenge overall, as it does for other NLP tasks with clinical documents.

Limitations to this study include a small number of participants with a limited number of total medications. There were more teen and female participants than adult and male participants and adults were a secondary target of the study. The text messages were not collected as part of a functioning medication management system and required prompting for patients to send messages. While the test dataset was not used for system training, the data were not blinded as all messages were reviewed prior to randomization, which could have influenced system development. However, we decided that as this is a new and emerging field of study, the need to review messages for “red flag” content that might indicate potential harm to participants or others outweighed the benefit of complete blinding.

CHAPTER V

FUTURE WORK

Introduction

These studies describe the development of a novel system for tagging medication concepts and action concepts in patient-generated medication messages. This is a new domain of NLP research with unique challenges. Future developments will be needed in order to fully implement PASTE as a generalizable toolkit for extracting medication information from patient SMS text messages.

Temporal Concepts

The knowledge required to record a medication administration event or other proposed medication management system function (e.g., cancel a medication reminder or record that a dose was skipped) include the medication concept, the desired action, and sometimes the time of the desired action. Finding and tagging of temporal concepts has been demonstrated in other NLP systems, typically using regular expressions (33, 52, 73-75). Adding this functionality to PASTE would ensure that all required knowledge necessary for recording medication actions by a medication management system would be extracted, tagged, and reported to the medication management system. Furthermore, the ability to detect and report time concepts will help overcome two challenges that were recognized during the evaluation study. First, implied medication actions (medication messages without an explicitly stated action) might be more accurately interpreted. For example, ‘lisinopril at 6pm’ might mean that the patient administered the medication at 6pm but it might also mean that the patient would like to schedule a reminder to take lisinopril at 6pm. Based on the time that the message was sent, the message could indicate very different things: e.g., a message sent after 6pm more likely indicates the patient took the medicine at that time. Comparison of time concepts found in medication messages with the timestamp of the message could help disambiguate implied medication actions. Secondly, recognition of temporal relationships of text messages may be necessary for proper functionality of the system. In one example a patient sent three successive text messages referring to his/her medication:

1. 'I forgot to take my med last n'
2. 'That means I have to double up today.'
3. 'That will lead to my stomach hurting.'

Each of these messages was sent 30 seconds apart by the same patient. The medication concept from the first message is carried through to both successive messages, with the association of needing to double the dose in the second message and the hypothetical side effect in the third message. Utilization of the message timestamp would be the only way to regularize these messages.

Hypothetical Statements, Questions, and Negation

The example above also demonstrates the need to recognize hypothetical statements. Even if the three messages were regularized, message two should not be interpreted as a medication administration event, but rather a hypothetical administration of twice the dose of the patient's regular medication dose. Other NLP systems have successfully implemented strategies for detection of hypothetical statements as well as questions and negation (33, 76). The F-measures for these concepts were lower than all other concepts in the evaluation study. Performance will need to be improved prior to integration into a live system if the system were to be implemented without verification. However, the design of MMH in concert with PASTE provides a method for integration in a clinical environment.

Ambiguous phrases, such as the above example, could be sent back to the patient for verification of the interpretation.

Other Features and Techniques

In a medication management system like MMH, patients might want to report more than administration of medications. For example, they may send messages describing qualitative associations such as adverse reactions, symptoms, or activities of daily living. Future development of PASTE will include expanded support for various medication management functions (e.g., refilling a medication, changing a medication dose, starting a new medication, finishing a medication). Additionally, improving disambiguation through updated and expanded lexicons, broadened semantic tagging, and contextual reasoning will be necessary. A natural progression for the system would be to extract and tag qualitative medication associations such as medication side-effects, activities of daily living, or patient opinions regarding medications – all data that have not historically been easily recorded or integrated into the medical record. Future technical improvements to PASTE could include incorporation of medication/action proximity or co-occurrence to improve disambiguation, the use of algorithmic scoring or prescription frequency weighting to improve candidate medication matching, incorporation of patient medication lists or recent electronic prescription data into medication messages to improve disambiguation of candidate medications, and integration of machine learning approaches.

CHAPTER VI

CONCLUSION

PASTE is an early step in bidirectional mobile communication as part of a medication management system. Evaluation of the system using patient-generated SMS text messages demonstrated that PASTE accurately extracts and tags medication concepts and desired medication actions with over 90% F-measure. This technology has the potential to improve guideline-based care, intercept drug interactions, stop unintentional medication overdoses, prevent improper scheduling of medications, and to gather real-time data about symptoms, outcomes, and activities of daily living. Integration into a medication management system has already begun. Future evaluation will be required to determine if PASTE integration supports dynamic, interactive patient-centered medication management and improved patient care.

APPENDIX A

TEEN TEXT MESSAGE QUESTIONNAIRE FOR PARENTS

Subject ID Number:

Date:

Adolescent LAST Name:

Adolescent FIRST Name:

Parent LAST Name:

Parent FIRST Name:

Who does the teenager live with? (Check all that apply.)

- ☐ Father
- ☐ Mother
- ☐ Grandparent
- ☐ Legal Guardian
- ☐ Other

What is your child's date of birth (mm/dd/yyyy)?

What is your child's gender?

- ☐ Male
- ☐ Female

Child's grade in school:

- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10
- ☐ 11

☐ 12

What type of school does your child currently attend?

☐ Public
☐ Private
☐ Home-schooled

What is your family's average income per year?

☐ Less than $20,000
☐ $20,001 – $40,000
☐ $40,001 – $70,000
☐ More than $70,000
☐ Decline to Answer

What is the highest grade in school that YOU completed?

☐ Grade 1-6
☐ Grade 7-11
☐ Grade 12 (High School Graduate)
☐ Some college, no degree
☐ College Degree
☐ Graduate school / Professional degree

What is your child's race?

☐ African American
☐ Asian
☐ Hispanic
☐ Native American
☐ Pacific Islander
☐ White/Caucasian

Adolescent email address:

Parent email address:

Teen's 10-digit CELL PHONE number (area code and cell phone number):

Your 10-digit CELL PHONE number (area code and cell phone number):

What cell phone platform is your child’s phone?

☐ Android
☐ iOS (Apple)
☐ RIM (Blackberry)
☐ Symbian (Nokia)

☐ Windows Mobile

☐ Other / Unknown

The cell phone manufacturer of your child's phone:

What type of text entry is available on your child's phone (check all that apply):

☐ Number Key Pad (multi-letters/key)

☐ Keyboard (1 letter/key)

☐ Touch Screen

☐ Don't Know

On average, how many text messages does your child send per month?

☐ Less than 100

☐ 100 – 499

☐ 500 – 999

☐ 1000 – 1499

☐ 1500 – 1999

☐ 2000 or more

How many different medications is your child scheduled to take regularly (more than 3 times per week)?

☐ 1

☐ 2

☐ 3

☐ 4

☐ 5 or more

Names and doses of your child's medications (if known):

REFERENCES

1. Sanders DL, Aronsky D. Biomedical informatics applications for asthma care: a systematic review. J Am Med Inform Assoc2006 Jul-Aug;13(4):418-27.

2. CTIA-The Wireless Association. CTIA Semi-Annual Wireless Industry Survey. [7/22/2011]; Available from: http://files.ctia.org/pdf/CTIA_Survey_Year_End_2010_Graphics.pdf.

3. Blumberg SJ LJ. Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January-June 2008. [17 Mar 2011]; Available from: http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200812.pdf.

4. Koivusilta LK, Lintonen TP, Rimpela AH. Orientations in adolescent use of information and communication technology: a digital divide by sociodemographic background, educational career, and health. Scand J Public Health2007;35(1):95-103.

5. Young S. What digital divide? Hispanics, African-Americans are quick to adopt wireless technology. Wall Street Journal Classroom Edition. [17 Mar 2011]; Available from: http://wsjclassroomedition.com/archive/06jan/tech_minoritywireless.htm.

6. CTIA - The Wireless Association. 2010 [7/22/2011]; Available from: http://www.ctia.org/advocacy/research/index.cfm/aid/10323.

7. Krishna S, Boren SA, Balas EA. Healthcare via cell phones: a systematic review. Telemed J E Health2009 Apr;15(3):231-40.

8. Ostojic V, Cvoriscec B, Ostojic SB, Reznikoff D, Stipic-Markovic A, Tudjman Z. Improving asthma control through telemedicine: a study of short-message service. Telemed J E Health2005 Feb;11(1):28-35.

9. Slagle JM, Gordon JS, Harris CE, Davison CL, Culpepper DK, Scott P, et al. MyMediHealth - designing a next generation system for child-centered medication management. J Biomed Inform Oct;43(5 Suppl):S27-31.

10. Saini SD, Schoenfeld P, Kaulback K, Dubinsky MC. Effect of medication dosing frequency on adherence in chronic diseases. Am J Manag Care2009 Jun;15(6):e22-33.

11. Ou HT, Feldman SR, Balkrishnan R. Understanding and improving treatment adherence in pediatric patients. Semin Cutan Med Surg Jun;29(2):137-40.

12. Bubalo J, Clark RK, Jr., Jiing SS, Johnson NB, Miller KA, Clemens-Shipman CJ, et al. Medication adherence: pharmacist perspective. J Am Pharm Assoc (2003) May-Jun;50(3):394-406.

13. Haynes RB, Ackloo E, Sahota N, McDonald HP, Yao X. Interventions for enhancing medication adherence. Cochrane Database Syst Rev2008(2):CD000011.

14. Rand CS, Wise RA. Measuring adherence to asthma medication regimens. Am J Respir Crit Care Med1994 Feb;149(2 Pt 2):S69-76; discussion S7-8.

15. Garcia-Lizana F, Sarria-Santamera A. New technologies for chronic disease management and control: a systematic review. J Telemed Telecare2007;13(2):62-8.

16. Benhamou PY, Melki V, Boizel R, Perreal F, Quesada JL, Bessieres-Lacombe S, et al. One-year efficacy and safety of Web-based follow-up using cellular phone in type 1 diabetic patients under insulin pump therapy: the PumpNet study. Diabetes Metab2007 Jun;33(3):220-6.

17. Bos A, Hoogstraten J, Prahl-Andersen B. Failed appointments in an orthodontic clinic. Am J Orthod Dentofacial Orthop2005 Mar;127(3):355-7.

18. Bramley D, Riddell T, Whittaker R, Corbett T, Lin RB, Wills M, et al. Smoking cessation using mobile phone text messaging is as effective in Maori as non-Maori. N Z Med J2005 Jun 3;118(1216):U1494.

19. Franklin VL, Waller A, Pagliari C, Greene SA. A randomized controlled trial of Sweet Talk, a text-messaging system to support young people with diabetes. Diabet Med2006 Dec;23(12):1332-8.

20. Hurling R, Catt M, Boni MD, Fairley BW, Hurst T, Murray P, et al. Using internet and mobile phone technology to deliver an automated physical activity program: randomized controlled trial. J Med Internet Res2007;9(2):e7.

21. Kim HS, Jeong HS. A nurse short message service by cellular phone in type-2 diabetic patients for six months. J Clin Nurs2007 Jun;16(6):1082-7.

22. Kim HS. A randomized controlled trial of a nurse short-message service by cellular phone for people with diabetes. Int J Nurs Stud2007 Jul;44(5):687-92.

23. Tang PC, Ash JS, Bates DW, Overhage JM, Sands DZ. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. J Am Med Inform Assoc2006 Mar-Apr;13(2):121-6.

24. Zeng K, Bodenreider O, Nelson SJ. Design and implementation of a personal medication record-MyMedicationList. AMIA Annu Symp Proc2008:844-8.

25. Britto MT, Wimberg J. Pediatric personal health records: current trends and key challenges. Pediatrics2009 Jan;123 Suppl 2:S97-9.

26. Kim MI, Johnson KB. Personal health records: evaluation of functionality and utility. J Am Med Inform Assoc2002 Mar-Apr;9(2):171-80.

27. Detmer D, Bloomrosen M, Raymond B, Tang P. Integrated personal health records: transformative tools for consumer-centric care. BMC Med Inform Decis Mak2008;8:45.

28. Koonce TY, Giuse DA, Beauregard JM, Giuse NB. Toward a more informed patient: bridging health care information through an interactive communication portal. J Med Libr Assoc2007 Jan;95(1):77-81.

29. Porter SC, Cai Z, Gribbons W, Goldmann DA, Kohane IS. The asthma kiosk: a patient-centered technology for collaborative decision support in the emergency department. J Am Med Inform Assoc2004 Nov-Dec;11(6):458-67.

30. Porter SC, Forbes P, Feldman HA, Goldmann DA. Impact of patient-centered decision support on quality of asthma care in the emergency department. Pediatrics2006 Jan;117(1):e33-42.

31. Brass EP. Consumer behavior in the setting of over-the-counter statin availability: lessons from the consumer use study of OTC Mevacor. Am J Cardiol2004 Nov 4;94(9A):22F-9F.

32. Hobbs JR. Information extraction from biomedical text. J Biomed Inform2002 Aug;35(4):260-4.

33. Harkema H, Dowling JN, Thornblade T, Chapman WW. ConText: an algorithm for determining negation, experiencer, and temporal status from clinical reports. J Biomed Inform2009 Oct;42(5):839-51.

34. Hahn U, Romacker M, Schulz S. MEDSYNDIKATE--a natural language system for the extraction of medical information from findings reports. Int J Med Inform2002 Dec 4;67(1-3):63-74.

35. Denny JC, Miller RA, Waitman LR, Arrieta MA, Peterson JF. Identifying QT prolongation from ECG impressions using a general-purpose Natural Language Processor. Int J Med Inform2009 Apr;78 Suppl 1:S34-42.

36. Denny JC, Smithers JD, Miller RA, Spickard A, 3rd. "Understanding" medical school curriculum content using KnowledgeMap. J Am Med Inform Assoc2003 Jul-Aug;10(4):351-62.

37. Fiszman M, Chapman WW, Aronsky D, Evans RS, Haug PJ. Automatic detection of acute bacterial pneumonia from chest X-ray reports. J Am Med Inform Assoc2000 Nov-Dec;7(6):593-604.

38. Fiszman M, Chapman WW, Evans SR, Haug PJ. Automatic identification of pneumonia related concepts on chest x-ray reports. Proc AMIA Symp1999:67-71.

39. Friedman C. Towards a comprehensive medical language processing system: methods and issues. Proc AMIA Annu Fall Symp1997:595-9.

40. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc1994 Mar-Apr;1(2):161-74.

41. Friedman C, Knirsch C, Shagina L, Hripcsak G. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. Proc AMIA Symp1999:256-60.

42. Haug PJ, Christensen L, Gundersen M, Clemons B, Koehler S, Bauer K. A natural language parsing system for encoding admitting diagnoses. Proc AMIA Annu Fall Symp1997:814-8.

43. 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 Care1995:284-8.

44. Hripcsak G, Austin JH, Alderson PO, Friedman C. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. Radiology2002 Jul;224(1):157-63.

45. Hripcsak G, Friedman C, Alderson PO, DuMouchel W, Johnson SB, Clayton PD. Unlocking clinical data from narrative reports: a study of natural language processing. Ann Intern Med1995 May 1;122(9):681-8.

46. Friedman C, Shagina L, Lussier Y, Hripcsak G. Automated encoding of clinical documents based on natural language processing. J Am Med Inform Assoc2004 Sep-Oct;11(5):392-402.

47. Bakken S, Hyun S, Friedman C, Johnson SB. ISO reference terminology models for nursing: applicability for natural language processing of nursing narratives. Int J Med Inform2005 Aug;74(7-8):615-22.

48. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp2001:17-21.

49. Elkin PL, Ruggieri AP, Brown SH, Buntrock J, Bauer BA, Wahner-Roedler D, et al. A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. Proc AMIA Symp2001:159-63.

50. Hersh WR, Donohoe LC. SAPHIRE International: a tool for cross-language information retrieval. Proc AMIA Symp1998:673-7.

51. Friedman C. A broad-coverage natural language processing system. Proc AMIA Symp2000:270-4.

52. Denny JC, Peterson JF, Choma NN, Xu H, Miller RA, Bastarache L, et al. Extracting timing and status descriptors for colonoscopy testing from electronic medical records. J Am Med Inform Assoc Jul-Aug;17(4):383-8.

53. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. J Am Med Inform Assoc Jan-Feb;17(1):19-24.

54. Aw A, Zhang M, Xiao J, Su J. A phrase-based statistical model for SMS text normalization. Proceedings of the COLING/ACL on Main conference poster sessions; Sydney, Australia: Association for Computational Linguistics; 2006.

55. Choudhury M, Saraf R, Jain V, Mukherjee A, Sarkar S, Basu A. Investigation and modeling of the structure of texting language. International Journal on Document Analysis and Recognition2007;10(3):157-74.

56. Kobus C, Fran\, \#231, Yvon o, G\, \#233, et al. Normalizing SMS: are two metaphors better than one? Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1; Manchester, United Kingdom: Association for Computational Linguistics; 2008.

57. Krawczyk S, Raghunathan K. Investigating SMS Text Normalization using Statistical Machine Translation. Stanford University, Stanford, CA http://nlpstanfordedu/courses/cs224n/2009/fp/27pdf2009.

58. Maier J, Ferens K, editors. Classification of english phrases and SMS text messages using Bayes and Support Vector Machine classifiers. Electrical and Computer Engineering, 2009 CCECE '09 Canadian Conference on; 2009 3-6 May 2009.

59. Xue Z, Yin D, Davidson B. Normalizing Microtext. Association for the Advancement of Artificial Intelligence http://wwwcselehighedu/~brian/pubs/2011/AAAI/normalizing-microtextpdf2011.

60. RxNorm. [7/22/2011]; Available from: http://www.nlm.nih.gov/research/umls/rxnorm/.

61. Tagg C. A corpus linguistics study of SMS text messaging. http://etheses.bham.ac.uk/253: University of Birmingham; 2009.

62. RxNormAPI. [7/22/2011]; Available from: http://rxnav.nlm.nih.gov/RxNormAPI.html.

63. Medicine NLo. RxNorm RESTful Web API. [17 Mar 2011]; Available from: http://rxnav.nlm.nih.gov/RxNormRestAPI.html.

64. Hripcsak G, Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. J Am Med Inform Assoc2005 May-Jun;12(3):296-8.

65. Evans DA, Brownlow ND, Hersh WR, Campbell EM. Automating concept identification in the electronic medical record: an experiment in extracting dosage information. Proc AMIA Annu Fall Symp1996:388-92.

66. Gold S, Elhadad N, Zhu X, Cimino JJ, Hripcsak G. Extracting structured medication event information from discharge summaries. AMIA Annu Symp Proc2008:237-41.

67. Levin MA, Krol M, Doshi AM, Reich DL. Extraction and mapping of drug names from free text to a standardized nomenclature. AMIA Annu Symp Proc2007:438-42.

68. SCOWL. [7/22/2011]; Available from: http://wordlist.sourceforge.net/.

69. PHP. [7/22/2011]; Available from: http://www.php.net/.

70. MySQL. [7/22/2011]; Available from: http://www.mysql.com/.

71. Facebook. [7/22/2011]; Available from: http://www.facebook.com/.

72. Twitter. [7/22/2011]; Available from: http://www.twitter.com/.

73. Zhou L, Hripcsak G. Temporal reasoning with medical data--A review with emphasis on medical natural language processing. Journal of Biomedical Informatics2007;40(2):183-202.

74. Zhou L, Friedman C, Parsons S, Hripcsak G. System architecture for temporal information extraction, representation and reasoning in clinical narrative reports. AMIA Annu Symp Proc2005:869-73.

75. Zhou L, Melton GB, Parsons S, Hripcsak G. A temporal constraint structure for extracting temporal information from clinical narrative. J Biomed Inform2006 Aug;39(4):424-39.

76. Chapman WW, Bridewell W, Hanbury P, Cooper GF, Buchanan BG. A Simple Algorithm for Identifying Negated Findings and Diseases in Discharge Summaries. Journal of Biomedical Informatics2001;34(5):301-10.