PASTE: patient-centered SMS text tagging in a medication management system

Shane P Stenner,[1,2] Kevin B Johnson,[1,3] Joshua C Denny[1,2]

[1]Department of Biomedical Informatics, Vanderbilt University, School of Medicine, Nashville, Tennessee, USA
[2]Department of Medicine, Vanderbilt University, School of Medicine, Nashville, Tennessee, USA
[3]Department of Pediatrics, Vanderbilt University, School of Medicine, Nashville, Tennessee, USA

Correspondence to
Dr Shane P Stenner, 2209 Garland Avenue, 400 Eskind Biomedical Library, Nashville, TN 37232-8340, USA; shane.stenner@vanderbilt.edu

Received 13 July 2011
Accepted 4 September 2011
Published Online First
8 October 2011

ABSTRACT

Objective To evaluate the performance of a system that extracts medication information and administration-related actions from patient short message service (SMS) messages.

Design Mobile technologies provide a platform for electronic patient-centered medication management. MyMediHealth (MMH) is a medication management system that includes a medication scheduler, a medication administration record, and a reminder engine that sends text messages to cell phones. The object of this work was to extend MMH to allow two-way interaction using mobile phone-based SMS technology. Unprompted text-message communication with patients using natural language could engage patients in their healthcare, but presents unique natural language processing challenges. The authors developed a new functional component of MMH, the Patient-centered Automated SMS Tagging Engine (PASTE). The PASTE web service uses natural language processing methods, custom lexicons, and existing knowledge sources to extract and tag medication information from patient text messages.

Measurements A pilot evaluation of PASTE was completed using 130 medication messages anonymously submitted by 16 volunteers via a website. System output was compared with manually tagged messages.

Results 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%, and the medication action term F-measure was 90%.

Conclusion Other studies have demonstrated systems that successfully extract medication information from clinical documents using semantic tagging, regular expression-based approaches, or a combination of both approaches. This evaluation demonstrates the feasibility of extracting medication information from patient-generated medication messages.

INTRODUCTION

Mobile technologies now 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 USA who send over 2.1 trillion text messages per year; almost every household in the USA has at least one cellular phone and over 26% are wireless-only households.[2] 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]

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 diabetic patients, stress levels, smoking quit rates, and patient self-efficacy.[6] Furthermore, cellular phone interventions have lowered missed appointment rates, decreased diagnosis and treatment times, and improved teaching and training of patients.[6] Ostojic *et al*[7] found that patients with asthma who received standard care plus peak expiratory flow monitoring and weekly treatment adjustment using text messaging for 4 months showed significantly greater improvement in asthma cough and night-time symptoms while lowering daily doses of inhaled corticosteroids and long-acting β-agonist than those who received only standard care.

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 short message service (SMS) text alerts.[6 7] 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 gather real-time data about symptoms, outcomes, and activities of daily living. However, a major challenge to a model for self-care utilizing bidirectional 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 an interoperable toolkit for extracting and tagging medication information from text messages.

Background

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.[8] A product of Project HealthDesign, MMH was funded through support from the Robert Wood Johnson Foundation 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 a new medication is being added or a medication reminder is being scheduled. Figure 2 shows the online medication list and a view of upcoming medication doses in MMH. The ongoing AHRQ-funded study is primarily concerned with improving adolescent patient adherence to scheduled daily preventive asthma medication (eg, inhaled corticosteroids). 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 (eg, '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 natural language processing (NLP) challenges. Patient messages cannot be expected to contain structured or complete medication information as one might find in a clinical document (eg, 'albuterol MDI 2 puffs inhaled'). Additionally, patients may communicate using medication names (brand or generic) or using nicknames for medications (eg, 'my puffer').

NLP systems identify probable structured concepts contained in narrative text. 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 KnowledgeMap.[9–19] Most common NLP systems use a process of sentence identification followed by syntactic and/or semantic parsing, followed by lexicon matching. 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.[20] 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. 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.[21 22]

Design objectives

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 all 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. Contextual ambiguity refers to instances where information can be understood in more than one way but the context of the information

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.

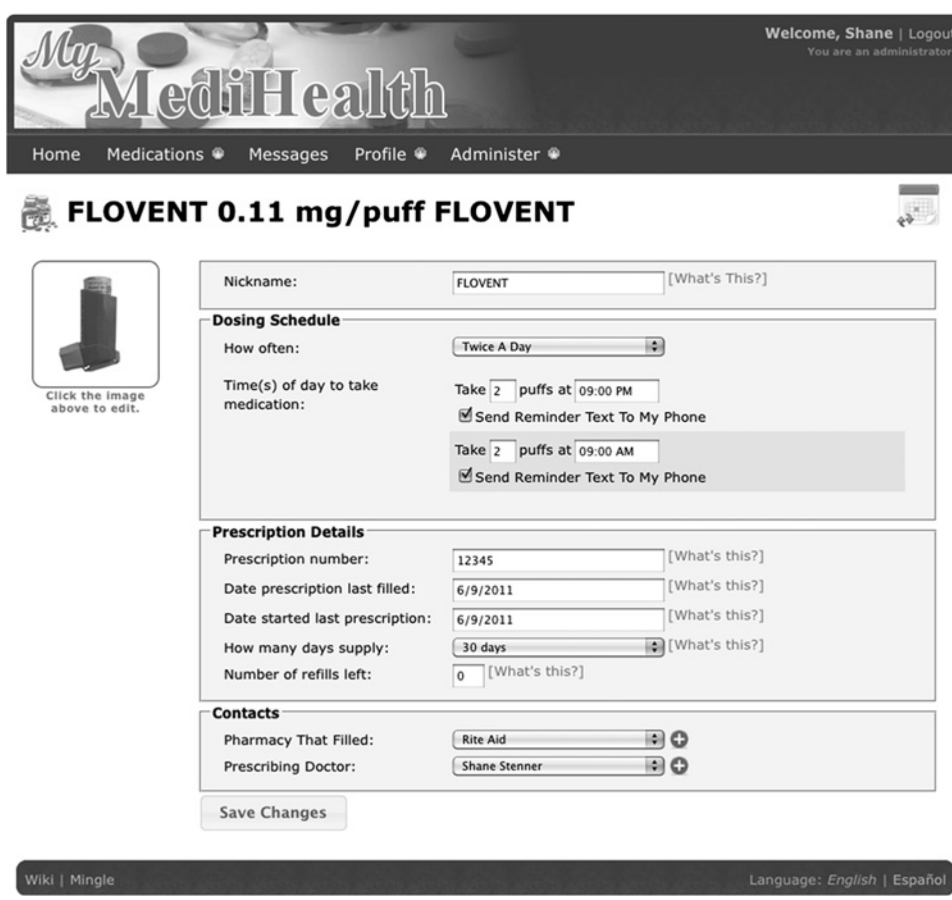

Figure 2 MyMediHealth screenshot showing medication list and upcoming doses. Patients can view their online list of medications, including images of their medications, and see various calendar views of upcoming medication doses.

may help resolve the ambiguity. For instance, '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 it 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) to take Advair. In a medication management system such as 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. 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. We envision that PASTE will provide structured medication information to MMH and other medication management systems that will oversee patient discourse for verification of medication names and actions. In this way, PASTE will have no direct patient interaction. Constraining PASTE to processing messages and providing structured output separates discourse system behavior and allows individual systems to handle message routing and workflow as well as management of medication-specific knowledge.

System description

The PASTE web service uses NLP techniques, custom lexicons, and existing knowledge sources, such as the National Library of Medicine's RxNav web service, 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.'[23] 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 early 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 are collected electronically and directly transferred 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.

We define a 'medication concept' as a medication name (either generic or brand) that is found in RxNorm, a general medication term (eg, 'pill', 'med', 'inhaler'), or a candidate medication that is algorithmically matched to a drug concept in RxNorm but is not an exact match (eg, a misspelled medication name). We define 'action concept' as a medication-associated term (eg, '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 because of the lack of an easily accessible corpus of patient-generated medication-related text messages.

Preprocessing and synonym replacement

The goal of the preprocessing step in PASTE is to remove unnecessary whitespace and symbols in preparation for the lookup tagger. In the piloted version of PASTE, we did not spell-check words before proceeding with tagging. A custom lexicon of over 1200 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.[24] 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 used to first convert text messages into 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.

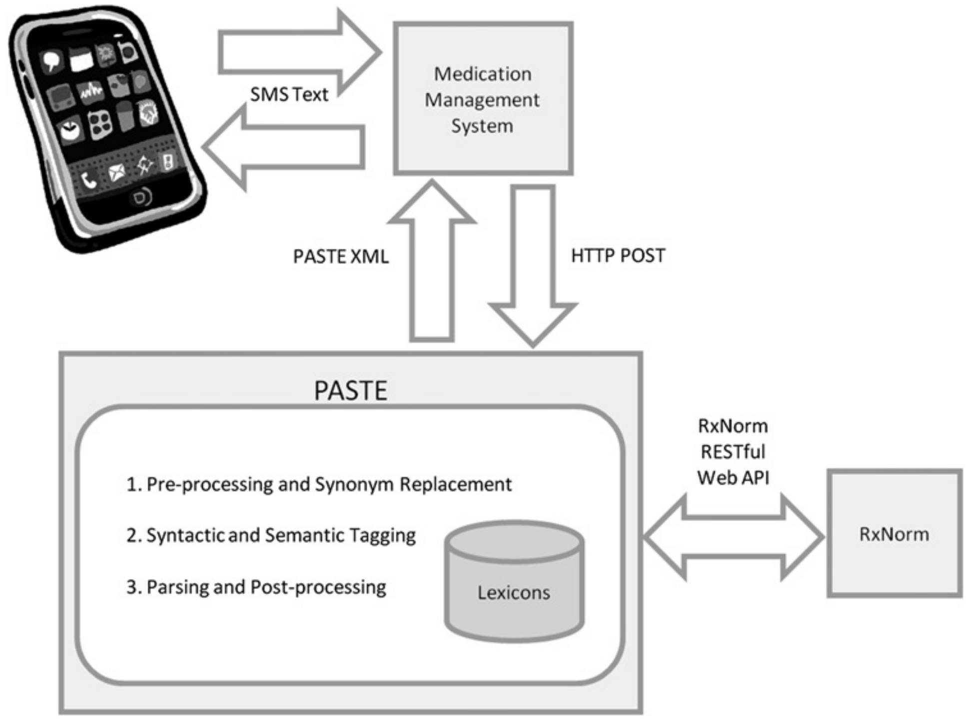

Figure 3 An overview of PASTE and the MyMediHealth systems. The medication management system, MyMediHealth, sends short message service (SMS) text message information to PASTE via the hypertext transfer protocol (HTTP) POST method. Structured extensible markup language (XML) is returned after processing by 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 use 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.[10]

A semantic tagging approach was used for finding medication information, including medication names, general medication terms, and medication-related actions. For medication names, we chose to use RxNav, a free web service hosted by the National Library of Medicine.[25] 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 rxcui number in the tag.[26] Data sent to RxNav do not contain, and cannot be directly linked to, any identifying patient information. A lexicon of general medication terms (eg, 'pill', 'antibiotic', 'meds') was constructed starting with medication forms (eg, '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 action 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 (eg, '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 completed, PASTE parses messages based on these unit boundaries, which we called 'segments.' We use regular expressions to find verified medications, actions, and negation signals (eg, 'didnt tk albuterol') in each segment.

'Conjunctive 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 (eg, 'advair', 'zyrtec', 'steroid') and an action term is tagged in the first two segments (eg, '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.

Pilot evaluation

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 anonymously submit sample medication messages via a website. We required that all 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

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

1. <u>Sample Input</u>
'used advair but 4got 2 take zyrtek & my steroid'

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

3. <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

4. <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>

5. <u>Syntactic and Semantic Tagging: Lookup untagged words to match 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>

6. <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>***

7. <u>Output</u>
```
<PASTE_XML>
        &lt;segment&gt;
        <action>administered</action>
        <medication>
              <candidate_med>advair</candidate_med>
              <verified>true</verified>
              <rxcui>301543</rxcui>
        </medication>
        &lt;/segment&gt;
        &lt;segment&gt;
        <action>snoozed</action>
        <medication>
              <candidate_med>zyrtek</candidate_med>
              <verified>false</verified>
              <suggestion>Zyrtec</suggestion>
              <rxcui>58930</rxcui>
        </medication>
        &lt;/segment&gt;
        &lt;segment&gt;
        <action>snoozed</action>
        <med_term>steroid</med_term>
        &lt;/segment&gt;
</PASTE_XML>
```

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 50 medication messages were used as the test set for evaluation. An internal medicine physician manually reviewed the 50 test set medication messages and annotated medication names, medication terms, and action terms. The same 50 messages were processed by PASTE and the structured output was manually compared with 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.[27] As a composite score of precision and recall, F-measure reflects the reliability or accuracy of the system's performance compared with the gold standard.

RESULTS

In total, there were 31 medication names and 21 medication terms identified by expert review. Ten medication messages did not contain an action, and three did not contain a medication concept (either a candidate medication or a general medication term). Only four messages contained more than one medication concept. Overall, there were 15 unique properly spelled medications ('verified'), six unique misspelled medications ('suggested'), and 11 unique 'medication terms' in the gold standard test set.

The results of the pilot test of PASTE, including precision, recall, and F-measure, are 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%.

DISCUSSION

In this paper, we present a novel patient medication SMS text message tagger, PASTE, and its pilot evaluation. The PASTE

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
Medical term	21	100	90	95
Action term	42	88	93	90
Negation	6	75	100	86
Conjunction	4	80	100	89

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,[28–30] 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 be 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 with a list of general English words such as the SCOWL list before searching for them in RxNorm.[20 31] We plan to make similar improvements to PASTE. In the future, the design goals of MMH would be to allow patients to enter user-specified nicknames (eg, '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. Additionally, MMH will provide PASTE with the patient's current medication list, which could help resolve misspelled medication names by providing candidate medications to match with the suggested medication list from RxNorm. These features, 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 (eg, '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 (ie, 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, (eg, '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 that the users entered a wide variety of messages—only two messages contained the same medication/action pair and similar syntax as another message. Another limitation is a single reviewer validated the output from PASTE. Another study is currently underway to collect medication messages from adolescent patients via SMS text message. 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. Future work should integrate such systems into live workflows, providing the possibility to create dynamic, interactive patient-centered medication management systems that may improve care.

Acknowledgments The authors would like to thank Nancy M Lorenzi, S Trent Rosenbloom, and Stuart T Weinberg for their critical reading of and suggestions for this article. None of the above received compensation for their contributions.

Funding SPS was funded by grant #5 T15 LM007450-09 from the National Library of Medicine.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Sanders DL,** Aronsky D. Biomedical informatics applications for asthma care: a systematic review. *J Am Med Inform Assoc* 2006;**13**:418–27.
2. *Ctia - The Wireless Association*. 2010. http://www.ctia.org/advocacy/research/index.cfm/aid/10323 (accessed 22 Jul 2011).
3. **Blumberg SJ,** Luke VJ. *Wireless Substitution: Early Release of Estimates From the National Health Interview Survey, January-June 2008*. http://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless200812.pdf (accessed 17 Mar 2011).
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 Health* 2007;**35**:95–103.

5. **Young S.** What digital divide? Hispanics, African-Americans are quick to adopt wireless technology. *Wall Street Journal Classroom Edition*. http://wsjclassroomedition.com/archive/06jan/tech_minoritywireless.htm (accessed 17 Mar 2011).
6. **Krishna S,** Boren SA, Balas EA. Healthcare via cell phones: a systematic review. *Telemed J E Health* 2009;**15**:231—40.
7. **Ostojic V,** Cvoriscec B, Ostojic SB, *et al*. Improving asthma control through telemedicine: a study of short-message service. *Telemed J E Health* 2005;**11**:28—35.
8. **Slagle JM,** Gordon JS, Harris CE, *et al*. MyMediHealth - designing a next generation system for child-centered medication management. *J Biomed Inform* 2010;**43**(5 Suppl):S27—31.
9. **Denny JC,** Miller RA, Waitman LR, *et al*. Identifying QT prolongation from ECG impressions using a general-purpose Natural Language Processor. *Int J Med Inform* 2009;**78**(Suppl 1):S34—42.
10. **Denny JC,** Smithers JD, Miller RA, *et al*. "Understanding" medical school curriculum content using KnowledgeMap. *J Am Med Inform Assoc* 2003;**10**:351—62.
11. **Fiszman M,** Chapman WW, Aronsky D, *et al*. Automatic detection of acute bacterial pneumonia from chest X-ray reports. *J Am Med Inform Assoc* 2000;**7**:593—604.
12. **Fiszman M,** Chapman WW, Evans SR, *et al*. Automatic identification of pneumonia related concepts on chest x-ray reports. *Proc AMIA Symp* 1999:67—71.
13. **Friedman C.** Towards a comprehensive medical language processing system: methods and issues. *Proc AMIA Annu Fall Symp* 1997:595—9.
14. **Friedman C,** Alderson PO, Austin JH, *et al*. A general natural-language text processor for clinical radiology. *J Am Med Inform Assoc* 1994;**1**:161—74.
15. **Friedman C,** Knirsch C, Shagina L, *et al*. Automating a severity score guideline for community-acquired pneumonia employing medical language processing of discharge summaries. *Proc AMIA Symp* 1999:256—60.
16. **Haug PJ,** Christensen L, Gundersen M, *et al*. A natural language parsing system for encoding admitting diagnoses. *Proc AMIA Annu Fall Symp* 1997:814—18.
17. **Haug PJ,** Koehler S, Lau LM, *et al*. Experience with a mixed semantic/syntactic parser. *Proc Annu Symp Comput Appl Med Care* 1995:284—8.
18. **Hripcsak G,** Austin JH, Alderson PO, *et al*. Use of natural language processing to translate clinical information from a database of 889,921 chest radiographic reports. *Radiology* 2002;**224**:157—63.
19. **Hripcsak G,** Friedman C, Alderson PO, *et al*. Unlocking clinical data from narrative reports: a study of natural language processing. *Ann Intern Med* 1995; **122**:681—8.
20. **Xu H,** Stenner SP, Doan S, *et al*. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc* 2010;**17**:19—24.
21. *Investigating SMS Text Normalization using Statistical Machine Translation*. http://nlp.stanford.edu/courses/cs224n/2009/fp/27.pdf (accessed 17 Mar 2011).
22. **Aw A,** Zhang M, Xiao J, *et al*. 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.
23. **National Library of Medicine.** http://www.nlm.nih.gov/research/umls/rxnorm/ (accessed 3 Mar 2011).
24. **Tagg C.** *A Corpus Linguistics Study of SMS Text Messaging*. http://etheses.bham.ac.uk/253: University of Birmingham, 2009.
25. **National Library of Medicine.** http://rxnav.nlm.nih.gov/RxNormAPI.html (accessed 17 Mar 2011).
26. **National Library of Medicine.** *RxNorm RESTful Web API*. http://rxnav.nlm.nih.gov/RxNormRestAPI.html (accessed 17 Mar 2011).
27. **Hripcsak G,** Rothschild AS. Agreement, the f-measure, and reliability in information retrieval. *J Am Med Inform Assoc* 2005;**12**:296—8.
28. **Evans DA,** Brownlow ND, Hersh WR, *et al*. Automating concept identification in the electronic medical record: an experiment in extracting dosage information. *Proc AMIA Annu Fall Symp* 1996:388—92.
29. **Gold S,** Elhadad N, Zhu X, *et al*. Extracting structured medication event information from discharge summaries. *AMIA Annu Symp Proc* 2008:237—41.
30. **Levin MA,** Krol M, Doshi AM, *et al*. Extraction and mapping of drug names from free text to a standardized nomenclature. *AMIA Annu Symp Proc* 2007:438—42.
31. *SCOWL*. http://wordlist.sourceforge.net/ (accessed 17 Mar 2011).