



ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Journal of Biomedical Informatics xxx (2003) xxx–xxx

Journal of  
Biomedical  
Informatics

www.elsevier.com/locate/yjbin

## Towards linking patients and clinical information: detecting UMLS concepts in e-mail<sup>☆</sup>

Patricia Flatley Brennan<sup>a,\*</sup> and Alan R. Aronson<sup>b</sup>

<sup>a</sup> University of Wisconsin-Madison, 372 Mechanical Engineering, 1513 University Avenue, Madison, WI 53706, USA

<sup>b</sup> National Library of Medicine, Building 38A, MS 54, 8600 Rockville Pike, Bethesda, MD 20894, USA

Received 18 September 2003

### 8 Abstract

9 The purpose of this project is to explore the feasibility of detecting terms within the electronic messages of patients that could be  
10 used to effectively search electronic knowledge resources and bring health information resources into the hands of patients. Our  
11 team is exploring the application of the natural language processing (NLP) tools built within the Lister Hill Center at the National  
12 Library of Medicine (NLM) to the challenge of detecting relevant concepts from the Unified Medical Language System (UMLS)  
13 within the free text of lay people's electronic messages (e-mail). We obtained a sample of electronic messages sent by patients  
14 participating in a randomized field evaluation of an internet-based home care support service to the project nurse, and we subjected  
15 elements of these messages to a series of analyses using several vocabularies from the UMLS Metathesaurus and the selected NLP  
16 tools. The nursing vocabularies provide an excellent starting point for this exercise because their domain encompasses patient's  
17 responses to health challenges. In successive runs we augmented six nursing vocabularies (NANDA Nursing Diagnosis, Nursing  
18 Interventions Classification, Nursing Outcomes Classification, Home Health Classification, Omaha System, and the Patient Care  
19 Data Set) with selected sets of clinical terminologies (International Classification of Primary Care; International Classification of  
20 Primary Care- American English; Micromedex DRUGDEX; National Drug Data File; Thesaurus of Psychological Terms; WHO  
21 Adverse Drug Reaction Terminology) and then additionally with either Medical Subject Heading (MeSH) or SNOMED Inter-  
22 national terms. The best performance was obtained when the nursing vocabularies were complemented with selected clinical ter-  
23 minologies. These findings have implications not only for facilitating lay people's access to electronic knowledge resources but may  
24 also be of assistance in developing new tools to aid in linking free text (e.g., clinical notes) to lexically complex knowledge resources  
25 such as those emerging from the Human Genome Project.

26 © 2003 Published by Elsevier Inc.

### 28 1. Introduction

29 Understanding and fulfilling patient information  
30 needs is a primary nursing responsibility most often met  
31 through careful exploration of the person's chief con-  
32 cerns, readiness to learn, and ability to process complex  
33 content, and verbal presentation of relevant content.  
34 Advances in health informatics, including the develop-

ment of standardized languages for health care, creation 35  
of internet-accessible knowledge resources and the ap- 36  
plication of electronic communication to the nurse-pa- 37  
tient relationship, afford nurses new opportunities for 38  
facilitating patient access to health information. Search 39  
engines permit anyone with a computer network con- 40  
nection and a web browser to selectively explore the vast 41  
resources of literature databases, full-text journals, and 42  
informative consumer health information on the inter- 43  
net. However, effective use of search engines places 44  
enormous cognitive demands on patients to identify 45  
appropriate terms for exploring complex and potentially 46  
unfamiliar knowledge resources. Judicious application 47  
of existing professional vocabularies and the natural 48  
language processing (NLP) tools used to manage them 49  
may open up an alternative approach to meeting patient 50

<sup>☆</sup> Support for this work was provided by the National Library of Medicine (LM 6247), the University of Wisconsin-Madison Graduate School (Kellet Professorship) and the Lister Hill Center, National Library of Medicine, National Institutes of Health.

\* Corresponding author. Fax: 1-608-263-5252.

E-mail addresses: klludwig@facstaff.wise.edu, pbrennan@engr.wisc.edu (P.F. Brennan), alan@nlm.nih.gov (A.R. Aronson).

51 information needs, one that capitalizes on the wide-  
52 spread acceptability of e-mail and other electronic  
53 communication mechanisms as a starting point for lo-  
54 cating and retrieving relevant health information. The  
55 purpose of this project is to explore the feasibility of  
56 detecting terms within the electronic messages of pa-  
57 tients that could be used to effectively search electronic  
58 knowledge resources and bring health information re-  
59 sources into the hands of patients.

60 As part of a larger program of work at the National  
61 Library of Medicine's Lister Hill Center, investigators  
62 develop and apply NLP approaches to analyze a wide  
63 range of text-based health science knowledge resources.  
64 For instance, the Indexing Initiative project investigates  
65 fully and semi-automatic methods for indexing the  
66 biomedical literature [1]. One Indexing Initiative ap-  
67 proach employs a concept-based indexing method based  
68 on MetaMap which maps citation text to concepts in the  
69 UMLS Metathesaurus [2]. Because MetaMap employs  
70 general-purpose NLP tools, and because source vocabu-  
71 laries of the UMLS Metathesaurus serve as index  
72 terms for a wide range of electronic health resources of  
73 interest to laypersons (professional literature, lay liter-  
74 ature, clinical records), these tools might be of particular  
75 value in helping identify salient concepts that appear in  
76 lay people's electronic communications. These concepts  
77 in turn could become the starting point of search and  
78 retrieval processes that directly link colloquial com-  
79 ments by patients to electronic knowledge resources,  
80 thus facilitating patient access to health information.

## 81 2. Background

82 Work in consumer health informatics over the past 15  
83 years has focused largely on creating internet-based  
84 health information resources, communication utilities,  
85 and patient-accessible clinical records systems. Insuring  
86 clear, accurate recognition of laypersons' information  
87 needs and presenting comprehensible, appropriate re-  
88 sponses to them form the core elements required for  
89 effective use of these resources. While most of the con-  
90 sumer health informatics efforts have focused on the  
91 *presentation* of health information to lay people, a  
92 smaller but critical part of these efforts attempt to *un-*  
93 *derstand* patients' expressions of health care concerns in  
94 a manner that enhances their ability to access electronic  
95 knowledge resources and clinical care record systems.

96 Precise understanding of the concerns raised by lay  
97 people in these electronic messages usually requires in-  
98 terpretation by a skilled health professional. Widman  
99 and Tong [3] examined 70 messages posted to a physi-  
100 cian-created website and determined the domain of  
101 clinical concerns (cardiology in 67 cases (96%) and re-  
102 lated conditions and procedures in 52 cases (74%)) as  
103 well as the goals of the inquiries (diagnosis (15), therapy

(48), prognosis (1), and patient education (6)). Widman  
and Tong also determined that psychological reassur-  
ance was a predominant reaction among those con-  
tacted afterwards. These concerns parallel those found  
during professional review of electronic messages re-  
ceived by a pediatric practice by D'Allessandro et al. [4].  
They cautioned professionals responding to electronic  
messages to avoid personalized responses but rather  
refer individuals to digital library resources. Such re-  
ferrals could benefit from the application of automatic  
indexing and mapping tools, such as those under de-  
velopment in the NLM's Indexing Initiative. Smith et al.  
[5] used one of these tools (MetaMap) to analyze human  
curators' culling of 504 terms reflecting findings and  
features from 109 questions posted to an electronic  
cancer information service. They found a match to  
UMLS concepts for over 95% of the abstracted words  
and phrases. Emerging NLP strategies that enhance  
automatic indexing may, in some cases, make it possible  
to eliminate the costly and time-consuming human cu-  
rator participation.

Some investigators advocate creating terminology  
tools, including a controlled consumer vocabulary, to  
enhance the articulation between the phrases employed  
by lay people and the standardized terminologies used  
by health professionals. Zeng et al. [6] evaluated the  
terms used by lay people to search a hospital's web site  
and found very poor matches between the terms used by  
these people and those found in the hospital's website.  
Rather than creating consumer vocabularies, they rec-  
ommended the development of extensive terminology  
support tools for clarifying lexical phrasing, semantic  
meaning, and users' mental maps during the search  
process. Conversely, Patrick et al. [7] determined that  
combining vernacular extensions to the UMLS facili-  
tates the expression of consumer health information  
needs during the search process. However, strategies  
that emphasize creating a patient/consumer terminology  
then mapping from that terminology to some recognized  
one are labor intensive and inconsistent with the current  
approaches that propose linking terms to referent ter-  
minology models.

Direct searching of selected internet sites and un-  
constrained free-text searching of the Web generally  
constitutes the most frequent occasions for lay people to  
undertake free-text entry of terms reflecting health  
concepts. Exploration of this behavior provides some  
insights regarding how lay people best be linked to those  
resources that are indexed using controlled, professional  
vocabularies. McCray et al. [8] describe their meticulous  
approach to insuring that search terms entered by pa-  
tients exploring the resources of ClinicalTrials.gov are  
recognized and result in the retrieval of relevant docu-  
ments. McCray's work details management of common  
problems of consumer expression, such as word variants  
and misspellings. Eysenbach and Kohler [9] evaluated

160 the effectiveness of internet-wide searching by lay peo-  
 161 ple. Although their sources lay people frequently use  
 162 search strategies that are suboptimal, most individuals  
 163 are satisfied with their results. This satisfaction appears  
 164 to result from the perceived ability by the individual to  
 165 compare and contrast information from sites they deem  
 166 relevant and of sufficient quality, and to reject those  
 167 found to be irrelevant or suspicious. When approaching  
 168 a search event, it is plausible that the individual user  
 169 defines the task as a query and deliberately selects a  
 170 small number of what they perceive to be appropriate  
 171 search terms. However, using NLP tools to identify  
 172 terms in the electronic messages of lay people may be  
 173 advantageous because it may help extract potential  
 174 search terms from a communications modality more  
 175 robust and forgiving than a search term box, and makes  
 176 good use of the apparent willingness of lay people to  
 177 peruse suggested reference sites.

178 The UMLS contains six source vocabularies relevant  
 179 to the domain of nursing. All nursing vocabularies have  
 180 been vetted by the professional nursing association and  
 181 have been subjected to rigorous evaluation for domain  
 182 completeness and utility in clinical information systems  
 183 [10]. Together these six vocabulary systems cover the full  
 184 range of the phenomena of concern for nursing: patient  
 185 problems, nursing strategies designed to manage these  
 186 problems, and patient outcomes. They provide a reason-  
 187 able set of reference vocabularies against which we may  
 188 examine and identify the health concerns of lay people.

189 It is critical to remember that these vocabulary sys-  
 190 tems have particular goals and that their intent is largely  
 191 to describe the nursing process and document nursing  
 192 care, not to represent the concerns of lay people.  
 193 However, given that the role of nursing in society is to  
 194 diagnose and treat human response to disease and hu-  
 195 man development, these vocabularies address a partic-  
 196 ular part of the patient experience not addressed in other  
 197 health care vocabularies. Indeed, Zielstorff et al. [11]  
 198 observed that the nursing vocabularies covered concepts  
 199 not found in the other vocabularies included in the  
 200 UMLS. It is logical, therefore, to anticipate that the  
 201 terms present in these vocabularies should provide good  
 202 coverage for the concerns raised by patients in electronic  
 203 mail messages to the nurse, and may be more precise  
 204 and less pathology-focused than other terminologies  
 205 found within the UMLS.

### 206 3. Methods

207 The purpose of this study was to evaluate the appli-  
 208 cation of MetaMap for detecting within the electronic  
 209 messages of patients the presence of terms found in the  
 210 UMLS. This section details MetaMap, the source vo-  
 211 cabularies employed in the project, the stimulus text,  
 212 and the procedure followed to conduct the evaluation.

#### 3.1. MetaMap

213

The MetaMap Indexing method of discovering and  
 ranking UMLS concepts in free text consists of applying  
 the MetaMap program to a body of text then ordering  
 the resulting concepts using a ranking function [12]. The  
 MetaMap program, itself, accomplishes three key  
 functions:

1. Parse text of free-text messages from the stimulus text  
 into simple noun phrases using the Specialist minimal  
 commitment parser.
2. Variant generation and discovery of term candidates  
 from the source terminologies from the UMLS.
3. Retention of the concepts associated with the best  
 mapping terms from the candidate list using a scoring  
 mechanism that evaluates the fit of each term from  
 the source vocabulary to the original phrase from  
 the stimulus text.

We constrained the MetaMap program in a few  
 ways. While it is possible to apply MetaMap in two  
 modes (basic, which relies on precise matching of stim-  
 ulus text and source vocabulary terms, and aggressive,  
 which tolerates extraneous words), we selected the more  
 conservative basic mode to process the text. In all trials  
 we evoke MetaMap in the basic mode and employed  
 options to restrict derivational variation to noun-ad-  
 jective cases, to suppress abbreviation expansion, and to  
 ignore word order in the input text.

We conducted several trials applying MetaMap to the  
 free text found in the electronic messages of patients,  
 first using only the nursing vocabularies and on suc-  
 cessive trials adding additional vocabulary sets. For  
 each MetaMap trial, we extracted summary information  
 such as the number of utterances and phrases processed  
 and the number of candidate matching concepts and  
 final mappings for each phrase. We also obtained de-  
 tailed information for each phrase including concepts  
 matched and their unique identifiers (CUIs) and se-  
 mantic types, the specific vocabularies from which the  
 matched concepts were extracted and the MetaMap  
 score indicating how well the concepts matched the  
 phrases.

#### 3.2. Source vocabularies

254

In earlier work we employed the full UMLS vocab-  
 ulary system as source vocabularies for identification of  
 relevant terms within the electronic messages sent by  
 patients to a nurse. This process revealed that about half  
 of the terms parsed from patient electronic messages  
 could be matched to one or more terms within the  
 UMLS; however, the extent of errors, including both  
 false positives and inappropriate mappings, suggested  
 that greater precision could be obtained if a subset of  
 vocabularies were employed in the exercise. The context  
 of the messaging activity (home nursing support fol-

lowing discharge from surgery) might more precisely captured if the source vocabularies were restricted to those more directly relevant to home care and nursing care issues.

Therefore, in the present study we initially restricted the MetaMap application to the source vocabularies from the UMLS to those defined as nursing vocabularies and related terminologies (for example, the Thesaurus of Psychological Terms (PSY2001)).<sup>1</sup> Nursing vocabularies included: the North American Nursing Diagnosis Association taxonomy of nursing diagnoses [13] (NANDA99), Saba's Home Health Care Classification [14] (HHC), the Omaha System [15] (OMS), the Nursing Interventions Classification [16] (NIC 99), the Patient Care Data Set [17] (PCDS), and the Nursing Outcome Classification [18] (NOC99).

Patients experience problems and express concerns across the full spectrum of physical, psychological, and therapeutic dimensions of health and illness. Thus, for completeness, we included additional UMLS source vocabularies in our evaluation (see Table 1). These vocabularies were identified by members of our team knowledgeable about the structure of the vocabularies, the concept and semantic types included, and their organization within the Semantic Network [19]. Every concept within the UMLS is organized under one of 134 semantic types, which arrange similar terms and concepts into coherent sets (e.g., "Individual Behavior," "Signs and Symptoms"). The Semantic Network depicts relationships among semantic types in the UMLS, for example, "is a type of," "is a consequence of."

### 3.3. Stimulus text

For the purpose of this evaluation we obtained as source text the electronic messages sent by patients to a clinical nurse during a field evaluation of an internet-based home care post-discharge support service [20]. The HeartCare intervention provided persons recovering from Coronary Artery Bypass Graft (CABG) surgery with communication utilities and recovery coaching information accessible through a standard web browser. Following a protocol approved by the relevant Human Subjects' Committees, patients who were medically stable following the CABG procedure were randomly assigned to one of three conditions: the HeartCare intervention, audio-taped discharge coaching instructions, or usual care. Patients had access to the HeartCare intervention for a six-month period. In a 24-month randomized field investigation, patients with access to HeartCare recovered faster, and with fewer negative symptoms, than persons with access to a standardized

Table 1

Source vocabularies employed in the evaluation of patient electronic messages

North American Nursing Diagnosis Association taxonomy of nursing diagnoses (NANDA99)
Saba's Home Health Care Classification (HHC)
Omaha System (OMS)
Nursing Interventions Classification (NIC 99)
Patient Care Data Set (PCDS)
Nursing Outcome Classification (NOC99)
International Classification of Primary Care (ICPC2E)
International Classification of Primary Care- American English (ICPC2AE)
Micromedex DRUGDEX (MMX01)
National Drug Data File (NDDF01)
Thesaurus of Psychological Terms (PSY2001)
WHO Adverse Drug Reaction Terminology (WHO97)
Medical Subject Heading 2003 (MSH_2003)
SNOMED International Version 3.5 (SMNI98)

discharge training method [21]. The messages selected for the evaluation described here came from the "E-Mail the Nurse" segment of the HeartCare intervention.

Three hundred and twenty-five sequential messages sent from patients to the project nurse were culled and anonymized according to the guidelines advanced by Sweeney et al. [22] first all surnames were extracted and replaced with subject codes unique to the current investigation. Then, references appearing in the text addressing other participants in the study were replaced with unique codes for those other participants. Finally, all elements (identifiers, headers, and message text) from the 84 message and none of the text was eliminated. An exact replica of a sample message is displayed in Fig. 1, marked with the delimiters needed for the MetaMap process. Misspellings, spacing and punctuation are as they appear in the original message.

### 3.4. Procedure

Two hundred and forty-one messages were retained. These messages were parsed using Rindfleisch's under-specified syntactic parser [23]; and the data, which included delimited phrases organized within message headers and messages, were run through the MetaMap program. To prepare the source text for analysis we arbitrarily defined the "Subject" line of each message as a "Title" and the body of each message as an "Abstract." This decision permitted us to use existing NLP tools without extensive modification. We conducted four trials, applying MetaMap in the basic mode to the stimulus text with the source vocabularies organized in the following manner:

1. Nursing Only
2. NursingPlus (the six nursing vocabularies plus International Classification of Primary Care (ICPC2E); International Classification of Primary Care- American

<sup>1</sup> Information on source vocabularies and natural language tools included in this project can be obtained from the UMLS Knowledge Server, <http://umlsks.nlm.nih.gov/>, accessed June 7, 2003.

Subject: medications

Dear Connie, I've been out of the loop for a few weeks. I had a setback with the appearance of a blood clot 2 weeks ago and was back in the hospital for a week. I was released a week ago Friday and now am on several new medications. With all these new meds, I feel nauseous almost all the time and frequently dizzy. I have a visiting nurse coming to see me 3x a week, and she monitors my blood pressure, temperature and checks my legs for possible clots. But nothing seems to help the nauseous feeling and I have little appetite. The medication I am now taking are □ I suspect the Lasix may be the culprit, since had been on it a LONG time ago and it made me nauseous, but I don't know. Do I really need to be on all of these now? I take alot of them at the same time (meal time), but should I change this and stagger them? What order should I take them, or are there alternatives to this medication for now? Any advise you could give me before I go back to see my internist on Tuesday would be helpful, then I could discuss it with him again. I see the cardiologist on Thursday and hope to be cleared to start cardiac rehab after that. Right now, however, it is slow going and discouraging. Thanks, Bill

Fig. 1. Sample message.

351	English (ICPC2AE); Micromedex DRUGDEX	structive here is the change in the number of errors in	380
352	(MMX01); National Drug Data File (NDFD01); The-	matching that occurred during the mapping process. We	381
353	saurus of Psychological Terms (PSY2001); WHO Ad-	defined errors based upon review of the source text by	382
354	verse Drug Reaction Terminology (WHO97)).	the primary author. Three types of errors were found:	383
355	3. NursingPlus and Medical Subject Headings (MeSH)	1. Recognition	384
356	4. NursingPlus and SNOMED International 3.5 (SNMI)	1.1. Stimulus terms are parsed in an overly granular	385
357	We conducted our evaluation on two levels: first, we	manner	386
358	evaluated the entire stimulus text as a unit successively	1.1.1. <i>the words "feeling nauseous" are parsed as</i>	387
359	using MetaMap to apply the four vocabulary sets. Next	<i>two separate terms</i>	
360	we selected a single message and examined the results of	2. Inappropriate terms, concepts or semantic types se-	389
361	the four mapping trials for it.	lected from the source vocabularies	390
362	Quantitative measures of vocabulary coverage pro-	2.1. Concept mapping is nonsensical	391
363	vide necessary but not sufficient evidence for appraising	2.1.1. <i>The word "I've" is mapped to the term VAL</i>	392
364	the adequacy of a source vocabulary to capture the	<i>(Semantic Type = Disease or Syndrome)</i>	
365	terms employed by lay people in their electronic mes-	2.2. One or more of the Semantic types is inappropri-	394
366	sages. It is also instructive to examine the source text in	ate for the context	395
367	detail to explore the actual matches and to pay particu-	2.2.1. <i>The word "back" is mapped to the term</i>	396
368	lar attention to the types of mismatches (errors) that	<i>"back" but the semantic type indicated was</i>	
369	occur in the process. For this purpose, we selected a	<i>"Body Location or Region"</i>	
370	message with a sufficiently dense stimulus text (count of	3. Matching	399
371	words, diversity of topics) to illustrate as broad a range	3.1. Polysemy—word that has more than one meaning	400
372	as possible of terms. Table 2 summarizes the evaluation	3.1.1. <i>The word "monitor" can refer both to a</i>	401
373	of the message depicted in Fig. 1 across all four vo-	<i>Health Care Activity and a Medical Device</i>	
374	cabularies.	3.2. Semantic types hold radically different meanings	403
375	This single message yielded 174 phrases. Performance	in different vocabularies	404
376	of the vocabularies on the four trials paralleled that	3.2.1. <i>The word has more than one meaning, e.g.,</i>	405
377	observed in the trials of the entire source test in that the	<i>"right" within SNOMED has the Semantic</i>	
378	mean number of matches per phrase remains close to 1	<i>Type "Spatial Concept" but within MeSH</i>	
379	and increases as vocabularies are added. What is in-	<i>is a "Qualitative Concept"</i>	

Table 2

Analysis of a single message (174 phrases parsed)

	Nursing Only	NursingPlus	NursingPlus + MeSH	NursingPlus + SNOMED
Candidates concepts	15	54	85	114
Mapped concepts	13	42	57	70
Phrases w/ one or more maps	12	43	50	57
Mean concepts/phrase	1.08	.98	1.14	1.22
Errors	3	23	37	39

Table 3

Results of the four MetaMap mapping on the full stimulus text

	Nursing Only	NursingPlus	NursingPlus + MeSH	NursingPlus + SNOMED
Candidates concepts	1016	3734	5786	7366
Mapped concepts	948	3094	4439	5078
Phrases w/ one or more maps	871	2863	3995	4383
Mean concepts/phrase	1.09 (0.28)	1.08 (0.30)	1.11 (0.35)	1.16 (0.38)

#### 409 4. Results

410 The stimulus text yielded 241 messages. Thematic  
411 content of the messages addressed symptom manage-  
412 ment, activities of every day living, and logistics of study  
413 participation. Application of the parsing process yielded  
414 15,326 distinct phrases, i.e., terms. The MetaMap pro-  
415 gram nominated matches to these terms from the con-  
416 cepts in the source vocabularies in a string matching  
417 process (see [2] for details on this approach). It is pos-  
418 sible to have more than one term from the stimulus  
419 document matched to a single concept in the source  
420 vocabularies. The summary of results for all four trials is  
421 presented in Table 3.

422 Candidate concepts include all terms from any in-  
423 cluded source vocabulary deemed to be an eligible  
424 match to the parsed phrase (For the Nursing Only run,  
425 this number is 1016). Mapped concepts include only  
426 those retained by the MetaMap evaluation process, 948  
427 for the Nursing Only run. Matches were found for only  
428 871 of the total 15,326 parsed phrases. Thus, for the  
429 Nursing Only run, the mean number of matches per  
430 phrase was 1.09 (s.d.0.28). A mean number of matches  
431 per phrase close to 1 is desired, as this would indicate a  
432 precise and unambiguous fit of the concept to the  
433 phrase.

434 Table 4 summarizes the vocabularies, the number of  
435 concepts and semantic types in each vocabulary, and the  
436 coverage provided by each vocabulary for the entire  
437 stimulus text. The number of terms from the vocabu-  
438 laries used in the mapping exercises is generally greater  
439 than the total number of terms in the vocabulary itself,  
440 indicating that some vocabulary terms are matched to  
441 more than one stimulus phrase. Two nursing vocabu-  
442 laries, the Omaha System and the Nursing Outcomes  
443 Classification, provided the best source of concepts for  
444 matching to the stimulus text. In general, as the number  
445 and complexity of vocabularies included in each trial of  
446 the mapping exercise increased, the coverage of source  
447 text phrases also increased.

#### 448 4.1. Discussion

449 The application of the MetaMap process to decode  
450 UMLS concepts from the electronic messages of lay  
451 people yielded promising results. Each trial detected  
452 important concepts present in the messages, but each  
453 trial also exposed significant limitations. While the  
454 Nursing vocabularies alone provided the best ratio of  
455 terms mapped to errors (13/3), the addition of salient  
456 clinical terminologies (NursingPlus) yields the best re-  
457 sults, balancing the coverage of the terminologies (42/

Table 4

Description of the vocabularies and the number of times terms from that vocabulary were used in a match in each of the four trials

Vocabulary descriptions			# Matched terms			
Vocabulary	# Terms	Semantic types	Nursing Only	Nursing Plus	Nursing Plus + MeSH	Nursing Plus + SNOMED
NANDA	169	12	272	258	256	251
NIC	10187	21	29	29	28	26
NOC	3056	31	468	426	415	398
HHC	335	26	112	98	96	91
OMS	539	48	395	351	336	335
PCDS	2229	25	105	87	85	82
PSY 2001	7671	119		2747	2691	2641
WHO 97	3831	37		258	260	231
ICPC2AE	210	14		8	8	8
ICPC2E	3757	42		244	242	234
MMS 01	11536	36		204	161	182
NDDF01	20088	47		241	198	218
MESH	516793	134			3519	
NOMED	164179	131				4245

174) with the ratio of concepts mapped to errors generated (42/23). The addition of the vocabularies in NursingPlus appears to offer benefit both because of the increase in concept coverage as well as the increase in the number of Semantic Types included in these vocabularies.

The original motivation of this study was to explore ways to use the nursing vocabularies for interpreting the concerns of lay people. Our study then should be considered a success in that it demonstrates that the Nursing Only vocabularies do provide an accurate, if incomplete, representation of the terms patient use in their electronic mail messages. When deemed correct, the terms from the source vocabularies provided a reasonable fit to the phrase from the stimulus text. For example, the phrase "the nauseous feeling" mapped to the term "nausea" extracted from several vocabularies from the Nursing Outcomes Classification and the NANDA Nursing Diagnoses, and the phrase "little appetite" mapped to the NOC term "Appetite."

The study was narrow in scope, and in no way does this work represent an attempt to interpret free text messages of lay people or generate automatic responses to complex, unstructured queries. It is critical to remember that the goal was to simply determine if concepts the standardized vocabularies used for professional purposes could be found among the free text messages of lay people. This well-circumscribed goal was accomplished, and our results show that it is possible to detect concepts from standardized vocabularies, including the nursing vocabularies, in the free-text of lay people. However, these results also show that large amounts of the free text messages of lay people do not include concepts from the standardized vocabularies present in the UMLS and that the mapping of these source terms to the stimulus phrase remains imprecise at best.

It is appropriate to conclude that in subsequent studies, the NursingPlus vocabulary set may provide the best starting point. Several caveats are in order; however, first, one must be mindful of the expectation of vocabulary coverage available from a single vocabulary. Without considering errors, the best performing vocabulary still only provided matches for one third of the phrases identified. However, given the wide range of topics found in even a single electronic message, it is unlikely that as large a number of phrases parsed would be recognized by specialized terminologies as is found in evaluation of the controlled text of professionals. This result does suggest that future work applying the NLM NLP tools to the free text of lay persons may require modifications in the parsing process so that idioms and other colloquial written speech employed by lay people be appropriately recognized.

Second, it is important to conceptualize the process of decoding patient electronic messages as a multi-step process. This study addressed the first step, determining

whether it is possible to identify plausible concepts from the UMLS. Additional post-processing activities, such as selection of the best search terms from a list of mapped candidates, might be in order, and require the generation of rules based on additional information such as the identity of the user. It is possible that post-processing of the list of mapped concepts to generate a reduced list of potential search terms will compensate for some of the errors evidenced here.

The application of the NLP tools, with the vocabularies of the UMLS restricted to the nursing vocabularies and related terminologies, provided greater precision with fewer errors than did our earlier application of the tools using the full UMLS. Therefore, our approach offers support for the application of the tools to consumer health issues; but it recommends that a restricted set of vocabularies be used during the application of the tools to unstructured text.

This application of the NLM tools is non-standard, and represents an extension not envisioned by the developers. We chose to restrict our processing of electronic messages to the constraints imposed by the tools, including delegation of the subject line to the expected "Title" designation and the message body to the "Abstract." Both the titling and abstracting of professional articles follows expectations of systematic thinking that may not be present in the construction of electronic messages by lay people. The primary compromise here is the presumption of coherence between the two. Multiple themes are present in almost every electronic message, thus suggesting that electronic messages consist of lists of almost unrelated topics. However, because both the parser and the source vocabularies target parsing of terms from source text independent of context, this modification is logical and acceptable within bounds.

The tools show some promise in identifying relevant professional health care vocabularies in the terms found in the free text of lay people employed in electronic messages. Additional work is required to scale the procedure employed here for routine use in the analysis of electronic messages. Greater attention to mapping accuracy, error determination, and management of errors is needed before general-purpose use of this approach can be advocated. In summary, the development of tools that will assist health professionals and lay people in identifying the health concepts present in the free-text electronic messages of lay people holds great promise, but awaits future pre-processing and post-processing strategies.

#### 4.2. Future applications

Patients are more likely to use electronic resources recommended by their health professionals, and refinements in the procedures described in this study could facilitate health professionals' recommendations without a concomitant investment in human curation. While our

568 approach is designed to complement both human cu-  
569 ration and automatic interpretation, it does not replace  
570 either. Nonetheless, it offers significant advantages to  
571 both of these in that it could be invoked deliberately as a  
572 plug-in utility, allowing the sender to mark up and gain  
573 access to health information prior to professional in-  
574 terpretation, which may in turn provide answers to some  
575 questions immediately or result in more precise and in-  
576 formative messages to professionals, thus making better  
577 use of the scarce resource of health professionals.

578 The approach employed here to detect concepts from  
579 recognized source vocabularies in the free text of pa-  
580 tients may provide a model for other situations in which  
581 the need exists to create linkages between free text de-  
582 scriptions of clinical or biomedical phenomena and  
583 electronic knowledge bases indexed by specialized vo-  
584 cabularies. For example, this process may help detect in  
585 the clinical notes of primary care clinicians caring for  
586 patients with heritable diseases the presence of concepts  
587 that are also found in taxonomies of genetic diseases.

## 588 5. Conclusion

589 The NLP tools of the NLM show promise for their  
590 utility in identifying words and phrases in the free text of  
591 lay people. Greater precision and coverage, and fewer  
592 errors, were found through applying the MetaMap pro-  
593 cess employing the NursingPlus vocabularies. Thus, while  
594 it is possible to detect within the written text of lay people  
595 terms found in standardized nursing vocabularies, full  
596 utility awaits more efficient pre-processing, that insures  
597 detection of the perhaps-unusual expressions of lay peo-  
598 ple and post-processing, which will refine the precision of  
599 mapping concepts to the vernacular of lay people.

## 600 References

- 601 [1] <http://ii.nlm.nih.gov> Accessed April 7, 2003.  
602 [2] Aronson AR. Effective mapping of biomedical text to the UMLS  
603 Metathesaurus: the MetaMap program. Proc AMIA Symp  
604 2001:17-21.  
605 [3] Widman LE, Tong DA. Requests for medical advice from patients  
606 and families to health care providers who publish on the World  
607 Wide Web. Arch Intern Med 1997;157(2):209-12.  
608 [4] D'Alessandro DM, D'Alessandro MP, Colbert SI. A proposed  
609 solution for addressing the challenge of patient cries for help

- through an analysis of unsolicited electronic mail. Pediatrics 610  
2000;105(6):E74. 611  
[5] Smith CA, Stavri P, Chapman WW. In their own words? A 612  
terminological analysis of e-mails to a cancer information service. 613  
Proc AMIA Symp 2002:697-701. 614  
[6] Zeng Q, Kogan S, Ash N, Greenes RA, Boxwala AA. Charac- 615  
teristics of consumer technology for health information retrieval. 616  
Methods Inf Med 2002:289-98. 617  
[7] Patrick TB, Monga HK, Sievert ME, Houston Hall J, Longo DR. 618  
Evaluation of controlled vocabulary resources for development of 619  
a consumer entry vocabulary for diabetes. J Med Internet Res 620  
2001;3(3):E24. 621  
[8] McCray AT, Dorfman E, Ripple A, et al. Usability issues in 622  
developing a Web-based consumer health site. Proc AMIA Symp 623  
2000:556-60. 624  
[9] Eysenbach G, Kohler C. How do consumers search for and 625  
appraise health information on the world wide web? Qualitative 626  
study using focus groups, usability tests, and in-depth interviews. 627  
BMJ 324:573-7. 628  
[10] Bakken (Henry) S, Warren JJ, Lange L, Button P. A review of 629  
major nursing vocabularies and the extent to which they have the 630  
characteristics required for implementation in computer-based 631  
systems. JAMIA 1998;5(4):321-8. 632  
[11] Zielstorff RD, Cimino C, Barnett GO, et al. Representation of 633  
nursing terminology in the UMLS metathesaurus: a pilot study. 634  
Proc Annu Symp Comput Appl Med Care 1993. 635  
[12] Aronson AR, Olivier B, Chang F, et al. The NLM indexing 636  
initiative. Proc AMIA Symp 2000:17-21. 637  
[13] NANDA. Nursing diagnoses: definitions and classification 2001- 638  
2002. North American Nursing Diagnosis Association, Philadel- 639  
phia: 2001. 640  
[14] Saba VK. Home Health Care Classification (HHCC): Nursing 641  
Diagnosis and Nursing Intervention. HHCC Web site. Available 642  
from: [www.sabacare.com](http://www.sabacare.com), accessed April 7, 2003. 643  
[15] Martin KS, Scheet NJ. The Omaha system: applications for 644  
community health nursing. Philadelphia: WB Saunders; 1992. 645  
[16] McCloskey Joanne C, Bulechek Gloria M, editors. Nursing 646  
interventions classification (NIC): iowa intervention project. St. 647  
Louis (MO): Mosby Year Book; 1999. 648  
[17] Ozbolt JG, Russo M, Stultz MP. Validity and reliability of 649  
standard terms and codes for patient care data. Proc Annu Symp 650  
Comput Appl Med Care 1995:37-41. 651  
[18] Johnson M, Maas M. The nursing outcomes classification. J Nurs 652  
Care Qual 1998;12(5):9-20. 653  
[19] <http://umlsks.nlm.nih.gov>, accessed June 7, 2003. 654  
[20] Brennan PF, Moore SM, Bjornsdottir G, Jones J, Visovsky C, 655  
Rogers M. HeartCare: an Internet-based information and support 656  
system for patient home recovery after coronary artery bypass 657  
graft (CABG) surgery. J Adv Nurs 2001;35(5):699-708. 658  
[21] Moore SM, Brennan PF, O'Brien RA. HeartCare: Early Results. 659  
(Abstract) AHA Nov 01. 660  
[22] Sweeney L. Three computational systems for disclosing medical 661  
data in the year 1999. Medinfo 1998;9(Pt 2):1124-9. 662  
[23] Rindfleisch TC, Rajan JV, Hunter L. Extracting molecular binding 663  
relationships from biomedical text. Appl Nat Lang Process 664  
2000:188-95. 665