

# Enhancing Patient Readability of Discharge Summaries with Automatically Generated Hyperlinks

**Mehnaz Adnan, Jim Warren**

*University of Auckland, Morrin Road, Glen Innes, Auckland, New Zealand  
madn002@aucklanduni.ac.nz, jim@cs.auckland.ac.nz*

**Martin Orr**

*University of Auckland, Morrin Road, Glen Innes, Auckland, New Zealand  
Waitemata District Health Board, Auckland, New Zealand  
Martin.Orr@waitematadhb.govt.nz*

## Abstract

*Patients often have difficulty in understanding medical concepts and vocabulary in their Discharge Summaries. We explore automatic hyper-linking to online resources for difficult terms as a means of making the content more comprehensible for patients. We use the Consumer Health Vocabulary (CHV) as a resource for scoring the difficulty of terms and to provide the most consumer-friendly synonyms. We implement a term extraction component providing semantic annotation using the KIM Knowledge and Information Management Platform. We hyperlink these terms to pages indexed by MedLinePlus to provide consumer-friendly online explanations. A web interface allows for viewing annotated Discharge Summaries and browsing search results. In a preliminary evaluation, the system was used to annotate eight Clinical Management sections of Discharge Summaries. The automatic hyper-linking provides good precision in linking to topically-relevant pages indexed by MedLinePlus. Our approach shows promise as a technology to deploy in future portals where consumers view their Discharge Summaries online.*

## 1. Introduction

A Discharge Summary (DS) provides a coherent picture of patients' condition at the time of discharge and is usually created by a health professional [1] for a number of audiences including patients and their families. Health consumers play an important role in managing their post discharge care [2], therefore patients need more understandable information to improve self care and to be empowered as a partner in their post discharge care. DSs in the Auckland metropolitan area are authored online and transmitted to General Practice via HL7 messages. At present, the Waitemata District Health Board in Auckland, New Zealand, our partner in this study, provides a hard copy of DSs to their patients at the time of discharge. However, with widespread availability of Internet, many jurisdictions are rolling out online services for their patients; for example, Kaiser Permanente patients access a wide range of provider information and services online [3]. Hence, our research is conducted in the context of current DSs, which are transmitted electronically, but only to providers, with an eye to a future when these DSs may be viewed online by patients as well.

Studies [4, 5] show that vocabulary and medical terminology are the key factors that affects a patient's comprehension. Zeng et al [6] observe that lack of familiarity with medical vocabulary is a major problem for patients in accessing available information. DSs contain terminologies that are difficult to understand by an average consumer. Even for a highly educated consumer, it is difficult to understand medical jargon, such as "haemorrhagic growth noted" and "haemoptysis and acute confusion." Our previous work identifies a significant amount of difficult terminology and abbreviations in DS contents [7].

Some of the most recent research to improve consumer friendliness of health record information is reported by Barto et al [8] and Trietler et al [9]. Barto et al have used “infobuttons” to provide contextualized explanation and links to references and educational materials for Pap smear reports. Trietler et al employ text translation as a method to simplify the Electronic Health Record texts for consumers. In a further study, pictographs are augmented with discharge instructions to improve comprehension [10].

The Semantic Web [11] technology has been applied in medical domain to link and define patients’ data to assist health professionals in their patients’ information retrieval process [12, 13]. Kostkova et al [14] demonstrated the use of the Semantic Web to provide contextualized browsing in a health portals’ web pages. In an earlier study semantic annotation was used to link news broadcast with related online resources [15]. However, semantic annotation from medical reports’ text to improve consumer understanding and access to informational resources is still largely unexplored.

In light of vocabulary difficulty in DSs, semantic annotations and automatic linking to web contents may produce a system to empower consumer health knowledge and improve self-care. Preparing for a future context where patients can access their DSs online, this paper describes a prototype semantic annotation system that augments the DS text with links to consumer friendly terms and web resources to make content more comprehensible. Furthermore, it will analyse how the automatic hyper-linking improves the browsing capabilities of consumers for finding related health information.

## 2. Materials and Methods

### 2.1. Materials

We collected 200 de-identified randomly selected Electronic Discharge Summaries (EDSs) from the clinical data repository of Waitemata District Health Board which manages two public hospitals, North Shore Hospital and Waitakere Hospital (Auckland, New Zealand) with yearly presentations of around 43,000 and 24,000 patients, respectively. The sample data was collected from a total of 62,674 EDSs generated during the period of June 2007 to July 2008. We retrieved 50 EDSs each from Emergency, Medicine, Surgery and Older Adult Health Services departments. The EDSs contains sections including diagnoses, admission reason, clinical management, discharge medications, follow up, procedures, and relevant laboratory results, as well as an Advice to Patient section. Clinical Management is the largest free text section authored by a health professional which provides a summary of in-hospital patient care.

### 2.2. Methods

Semantic annotation is a process of assigning entities in the text to their semantic descriptions [11]. Automatic semantic annotations generates metadata for the terms or phrases contained in the background knowledge base. Semantic annotations can be used to provide automatic highlighting, hyper-linking and availability of relevant knowledge for the annotated named entities [16]. Automatic hyper-linking can provide dynamically selected web content and customizing the knowledge to better meet user requirements.

We employed the open-access collaborative (OAC) consumer health vocabulary (CHV) [4] for the purpose of semantic annotation in DS text. The CHV provides consumer-specific medical terms and concepts. Each CHV concept has a consumer friendly display name, also called the CHV preferred term. Each term in CHV has three associated familiarity scores: a *frequency-based term score* (calculated by a support-vector machine model based on term occurrence frequency in several health text corpora), a *context-based term score* (calculated based on term co-occurrence patterns in a health-specific query log data), and a *context-based concept score* (calculated on the basis of concept co-occurrences in medical literature and log data as well as semantic relations in medical vocabularies). The term scores reflect the string-level difficulty to estimate the likelihood the term will be recognized by an average consumer. The concept score estimates the concept-level difficulty for consumers. The scores range between 0 and 1, with a score of 0.8 to 1.0 representing “likely”, 0.5 to 0.8 “somewhat likely” and below 0.5 “not likely” for a term to be familiar to a consumer. Some CHV terms did not have scores assigned (indicated as a -1). To provide semantic annotations for the difficult terms in DS text, we use CHV terms having any one of their three CHV scores between 0 and 0.5.

Two strategies were employed to mitigate the vocabulary difficulty of free text in DSs: 1) synonym provision and 2) search and retrieval of web resources.

### **Synonym Provision**

Medical concepts often have alternate names (synonyms), one of which may be easier for the consumer to understand than others. Linking a term with synonyms and, more specifically, providing its more comprehensible synonym may improve comprehension. For example, “haemoptysis” is easier to understand if “coughing up blood” is also provided in parallel.

### **Search and Retrieval of Web Resources**

Consumers are commonly using the world wide web to seek medical and health information, but they often experience difficulty in finding specific health information online [17]. Keselman et al [18] asserts that increased understanding of medical text can be accomplished by facilitating “precise information retrieval” and with optimal result presentation. While not all terms have easy synonyms to be displayed, and consumers would like to seek more knowledge about the terminologies used in their DSs, it would be advantageous to provide dynamically selected customized web contents.

In the next section we introduce our semantic annotation approach to provide synonyms and hyperlinks to web resources in DS texts.

## **3. Incorporation of Semantic Annotations**

The text in DSs was semantically annotated using the Knowledge and Information Platform KIM [16]. KIM performs information extraction and semantic annotation for the named entities in text documents, thus allowing the CHV terms in DS text to be identified and annotated. We used KIM to parse the DS text and map terms to CHV concepts and their synonyms. As described in the previous section, all CHV concepts have pre-defined consumer friendly display names. This enables us to link CHV terms with their preferred display names and subsequently with their synonyms.

Furthermore, we use the query interface of MedlinePlus, a consumer health portal from the US National Library of Medicine, to provide automatic linking to related online resources. For each CHV term, its corresponding CHV preferred term was used in the search query to retrieve relevant web pages. The reason behind using the CHV preferred term in search query was that MedlinePlus has a built-in thesaurus which automatically adds the synonyms in the search procedure.

The KIM platform has a web user interface (KIM web UI) to view annotated text and its corresponding annotations. Figure 1 shows an example of DS with its Clinical Management section annotated<sup>1</sup> in the KIM web UI. The annotated CHV terms in the text are highlighted. Clicking on one of the highlighted term will bring up the full details of the corresponding annotations in the KIM Explorer as show in Figure 2. This includes synonyms for the annotated term with a title of “Synonym”. The phrase preceding “a CHV Preferred Term” shows the consumer friendly synonym of the annotated term. The last entry, having a title of “Search MedlinePlus”, provides a hyperlink to MedlinePlus search results for the annotated term. Clicking on the hyperlink link will open a new browser window (Figure 3) containing the MedlinePlus search results for the corresponding CHV preferred term.

---

<sup>1</sup> To preserve confidentiality of patient data, this illustrative Clinical Management section is taken from *National Discharge Summary: Data Content Specifications* published by the National E-Health Transition Authority, Australia.



Figure 1 - Screen Shot of an Annotated Clinical Management Section in the KIM Web UI

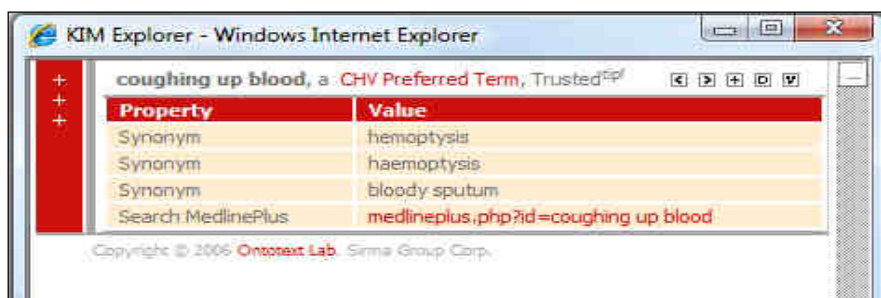


Figure 2 - KIM Explorer Appears after Clicking Annotated Term (haemoptysis)

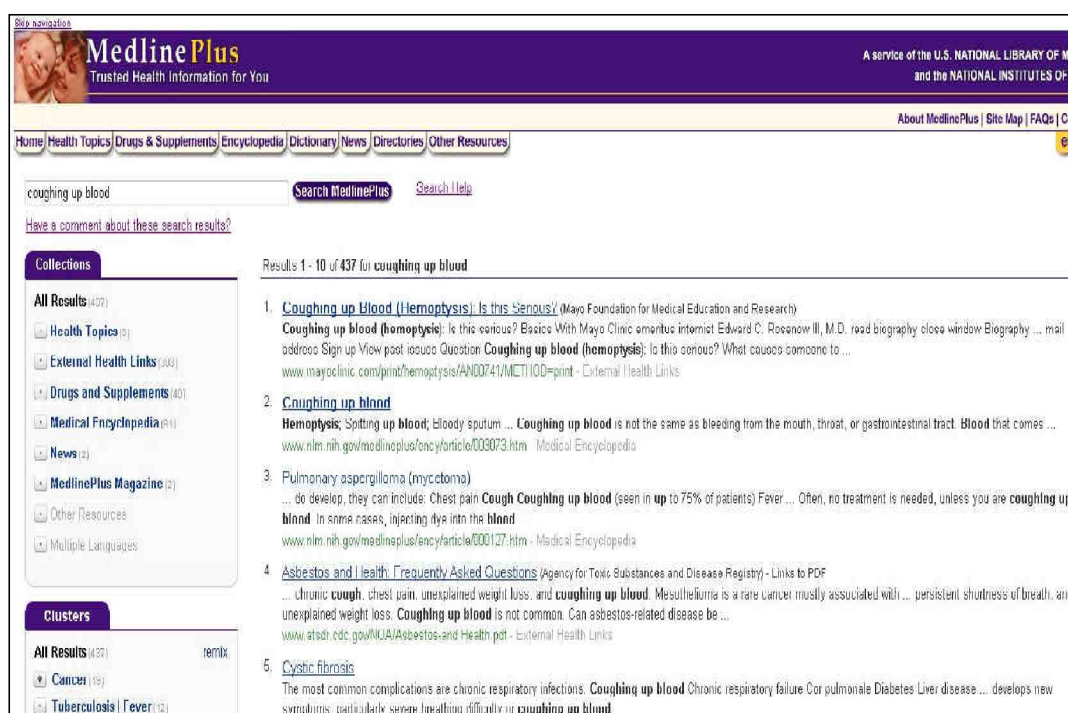


Figure 3 - MedlinePlus Search Results for the Annotated Term (haemoptysis)

**Table 1 - Cumulative Average Precision of 73 Search Results after  $k$  Web Pages Retrieved**

Number of Web Pages Retrieved ( $k$ )	Strict	Lenient
1	0.70	0.99
2	0.66	0.99
3	0.64	0.99
4	0.62	0.99
5	0.59	0.99

#### 4. Preliminary Evaluation

A preliminary evaluation of the system’s performance was conducted by annotating eight Clinical Management sections from of DS data (choosing two at median length for each speciality). We wanted to measure the utility of the automatic hyper-linking capability of the system for assisting consumers in retrieving related online resources. For this purpose, the top five MedlinePlus search results were indexed for each unique annotated term appearing in any of the eight Clinical Management sections. Then *precision*, a metric which reflects the share of relevant documents in the search results, was calculated for each of the five search results. The system identifies 117 CHV terms, out of which 112 are unique. For the purposes of measuring the relevance of retrieved web pages, non health specific terms (such as “history, unremarkable”) were not considered as MedlinePlus does not provide search results for these ‘daily-usage’ terms. This results in an analysis of search results conducted for 73 annotated terms. For each indexed web page, it was determined whether the searched term or its CHV preferred term is present in the page’s title or in the content. Our hypothesis is that if a term is in page title then the retrieved page provides “more specific information” about the searched term. On the other hand, if the term is in page content, than the web page is likely to provide “some information” about the searched term.

Table 1 shows the cumulative average precision after each of top five retrieved web pages under two conditions. In the first condition, *strict*, web page was only considered correct if the term is in the title of web page, but in the second, *lenient*, it was considered correct if a web page contains the term in its content. The results of the evaluation show that the system achieved very high average precision (99%) in retrieving web pages under the *lenient* condition. Although the performance is decreased by 29% to 40% under the *strict* condition, we see that 70% of the time the first link from MedlinePlus provides strict relevance. In addition, we have calculated that 79% of search results return at least one strictly relevant web page among the first five MedlinePlus results.

#### 5. Discussion

We have designed a prototype semantic annotation system to improve the readability of DSs for consumers. The system is designed to reduce the vocabulary difficulty by providing knowledge support through automated hyper-linking to online resources. The system can assist in answering questions occurring to a lay person while reading their own DS.

This study focused on providing automated links to related online resources to reduce the vocabulary difficulty in DSs. While text translation [9] and contextualized links [8] have been used to reduce the vocabulary difficulty and provide knowledge support in medical text, we explored a different approach: consumer friendly synonym provision and hyperlink generation to online resources. Text translations are useful, but this mechanism hides the original text from the consumer. Infobuttons are also helpful, however, they generate contextualized rather than term-based links and may not provide vocabulary and knowledge support for most of the difficult terms in the report.

In preliminary evaluation, 99% of web pages suggested for hyper-linking have the targeted term in their body, hence providing some related information to consumer, which is encouraging. However, when using the strict criterion that a term must appear in the title of the linked page, there is some potential to improve from the current mean of 59% precision. 79% of search results provide at least one

strictly relevant web page among the top 5 links; so we are exploring an extension of our methods to automatically select the most relevant links from among a set.

At present, the system is using MedlinePlus to provide hyper-inking from CHV terms in DS text to consumer specific online resources. However, the system could also be extended to incorporate clinical terms from UMLS or SNOMED. Moreover, resources such as PubMed or Cochrane could be used to render the same DSs for clinical users with hyperlink annotations more appropriate to their knowledge levels.

Our evaluation was preliminary and had limitations. One limitation is that we analyzed the retrieved web pages in a computational way by matching terms in web pages' title and body. While the relevance of hyperlinks based on CHV terms appears high, the CHV system is still growing and is not as complete as a system like SNOMED (consider "tachypnoea" in Figure 1, which clearly warrants a supporting link). A further study to evaluate the usefulness of the system in terms of actual consumer feedback is a further important step for this research program.

## 6. Conclusion

We have designed a prototype system that provides automated hyperlinks to relevant online information sources from DS text. Our work suggests that semantic annotation augmented with customized links to relevant online knowledge sources can aid consumer comprehension of DS text. While such an automated hyperlink capability has the potential to improve consumer readability, we also recognize the importance of eliminating redundant results and providing more related web resource to the consumer.

## 7. Acknowledgments

The authors gratefully acknowledge the help of the Waitemata District Health Board staff in making available the de-identified Discharge Summary data essential to this research. This work was supported by a Higher Education Commission, Pakistan scholarship. The research protocol was approved by the University of Auckland Human Participants Ethics Committee under protocol number 2008/221 and by the Waitemata District Health Board Knowledge Centre.

## 8. References

- [1] Barretto S, Chu S, Browne E and Clapton W. National Discharge Summary: Data Content Specifications Version 1.0 2006 [cited; Available from: [http://203.110.153.105/index.php?option=com\\_docman&task=doc\\_view&gid=175&Itemid=139](http://203.110.153.105/index.php?option=com_docman&task=doc_view&gid=175&Itemid=139)
- [2] Maloney LR and Weiss ME. Patients' perceptions of hospital discharge informational content. *Clin Nurs Res.* 2008 Aug;17(3):200-19.
- [3] Zhou YY, Garrido T, Chin HL, Wiesenthal AM and Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. *Am J Manag Care.* 2007 Jul;13(7):418-24.
- [4] Keselman A, Tse T, Crowell J, Browne A, Ngo L and Zeng Q. Assessing consumer health vocabulary familiarity: an exploratory study. *J Med Internet Res.* 2007;9(1):e5.
- [5] Clarke C, Friedman S, Shi K, Arenovich A and Culligan C. Emergency department discharge instructions comprehension and compliance study. *Cjem.* Jan 2005;7(1):7.
- [6] Zeng Q, Kogan S, Ash N and Greenes RA. Patient and clinician vocabulary: how different are they? *Stud Health Technol Inform.* 2001;84(Pt 1):399-403.
- [7] Adnan M, Warren J and Orr M. Assessing Text Characteristics of Electronic Discharge Summaries and their Implications for Patient Readability. 2009. Submitted in AMIA Annu Symp 2009.

- [8] Baorto DM and Cimino JJ. An "infobutton" for enabling patients to interpret on-line Pap smear reports. Proc AMIA Symp. 2000;Annual Symposium.:47-50.
- [9] Zeng-Treitler Q, Goryachev S, Kim H, Keselman A and Rosendale D. Making texts in electronic health records comprehensible to consumers: a prototype translator. AMIA Annu Symp Proc. 2007;Annual Symposium Proceedings/AMIA Symposium.:846-50.
- [10] Zeng-Treitler Q, Kim H and Hunter M. Improving Patient Comprehension and Recall of Discharge Instructions by Supplementing Free Texts with Pictographs. AMIA Annual Symposium; 2008; 2008.
- [11] Berners-Lee T, Hendler J and Lassila O. The Semantic Web. Scientific American 2001;284:36.
- [12] Zillner S, Hauer T, Rogulin D, Tsymbal A, Huber M and Solomonides T. Semantic Visualization of Patient Information. CBMS'08 2008; 2008. p. 296-301.
- [13] Berlanga R, Jimenez-Ruiz E, Nebot V, Manset D, Branson A, Hauer T, et al. Medical Data Integration and the Semantic Annotation of Medical Protocols Computer-Based Medical Systems; 2008; 2008. p. 644-9.
- [14] Kostkova P, Diallo G and Jawaheer G. User Profiling for Semantic Browsing in Medical Digital Libraries The Semantic Web: Research and Applications Springer Berlin / Heidelberg 2008:827-31
- [15] Dowman M, Tablan V, Cunningham H and Popov B. Web-assisted annotation, semantic indexing and search of television and radio news. 14th International World Wide Web Conference (WWW2005); May 2005; Chiba, Japan; May 2005. p. 225–34.
- [16] Kiryakov A, Popov B, Terziev I, Manov D and Ognyanoff D. Semantic annotation, indexing, and retrieval Web Semantics: Science, Services and Agents on the World Wide Web. 2004;2(1):30.
- [17] Keselman A, Browne AC and Kaufman DR. Consumer health information seeking as hypothesis testing. Journal of the American Medical Informatics Association. 2008 Jul-Aug;15(4):484-95.
- [18] Keselman A, Logan R, Smith CA, Leroy G and Zeng-Treitler Q. Developing informatics tools and strategies for consumer-centered health communication. Journal of the American Medical Informatics Association. 2008 Jul-Aug;15(4):473-83.