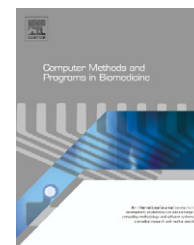




ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

MedicoPort: A medical search engine for all

Aysu Betin Can*, Nazife Baykal

METU Informatics Institute, İnönü Bulvarı, 06531 Ankara, Turkey

ARTICLE INFO

Article history:

Received 26 September 2006

Received in revised form

4 January 2007

Accepted 4 January 2007

Keywords:

Medical information retrieval

UMLS

Topical crawler

Internet

ABSTRACT

We present a new next generation domain search engine called MedicoPort. MedicoPort is a medical search engine designed for the users with no medical expertise. It is enhanced with the domain knowledge obtained from Unified Medical Language System (UMLS) to increase the effectiveness of the searches. The power of the system is based on the ability to understand the semantics of web pages and the user queries. MedicoPort transforms a keyword search into a conceptual search. Through our system we present a topical web crawling technique and indexing techniques empowered by the semantics information.

MedicoPort aims to generate maximum output with semantic value using minimum input from the user. Since MedicoPort is designed to help people seeking information about health on the web, our target users are not medical specialists who can effectively use the special jargon of medicine and access medical databases. Medical experts have the advantage of shrinking the answer set by expressing several terms using medical terminology. MedicoPort provides the same advantage to its users through the automated use of the medical domain knowledge in the background. The results of our experiments indicate that, expanding the queries with domain knowledge, such as using the synonyms and partially or contextually relevant terms from UMLS, increase dramatically the relevance of an answer set produced by MedicoPort and the number of retrieved web pages that are relevant to the user request.

© 2007 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Internet is a vast source of information from different domains. Among these different domains, health and medical information is one of the leading domains with huge amount of information. For example, MEDLINE reports containing approximately 13 million references to biomedicine articles.¹ Having such massive amount of information at disposal, people become more conscious about their health problems. Being more aware of their health, patients wish to be armed with knowledge about their medical problems. Nowadays, patients tend to prefer to be more informed before, during, and after consulting to their physicians. As opposed to the classical patient behavior, these patients wish to take part in the deci-

sion process for their health problems. For this purpose, such patients search the Internet, which is the easiest and fastest way to access information, and try to dig information out. This situation leads to a growing number of medical inquires on the Internet by the people with no medical training. Therefore, there is a growing need for a medical information retrieval techniques and tools to help ordinary people in searching health information on the Internet.

A recent experimental study [1] on how people consume health information on the Internet shows that people tend to start a medical information search on general search engines rather than medical libraries and medical societies. On the other hand, information retrieved from public institutions and scientific sites as well as links originated from these sites are

* Corresponding author. Tel.: +90 312 210 3740/3741; fax: +90 312 210 3745.

E-mail addresses: aysu@ii.metu.edu.tr (A.B. Can), baykal@ii.metu.edu.tr (N. Baykal).

¹ <http://www.nlm.nih.gov/pubs/factsheets/medline.html>.

0169-2607/\$ – see front matter © 2007 Elsevier Ireland Ltd. All rights reserved.

doi:10.1016/j.cmpb.2007.01.007

considered more trustworthy by the subjects. The study also concludes that the information with heavy medical terminology is not useful. For instance, some of the subjects continued their search even after finding a page with the answer since they did not understand the information presented on the page.

To address the problem of a medical search engine tailored for non-expert medical information seekers, we have developed the MedicoPort system. MedicoPort is a search engine built to retrieve medical information from the Internet. It is not for literature database search. Therefore, the answer set provided by MedicoPort is not limited to a specific medical information collection. MedicoPort is for anybody who wishes to perform a web search about health information. Our system has the ability to take a simple user query and match that query to a concept (or a set of concepts in a context) in medical terminology. The system ranks the query results based on the document relevance and the relevance of the links originated from that document. MedicoPort is empowered with a complete medical lexicon (Unified Medical Language System [2,3] Specialist Lexicon) and the Unified Medical Language System (UMLS) Metathesaurus. By exploiting this lexicon and metathesaurus, MedicoPort filters irrelevant web pages in addition to relieving its user from medical terminology.

In this paper we present the MedicoPort system including its role, structure and capabilities. The contributions of this paper can be summarized as follows:

- A new search engine system to retrieve medical information available on the Internet, not on a specific medical literature database, to the people with no medical training.
- A system that is able to take a simple specification of the non-medical user and match the specification to a concept or a set of concepts of a context in medical terminology through the use of a complete medical lexicon and a metathesaurus.
- A topical web crawler, called Lokman, and indexing strategy that exploits the advantage of the medical domain knowledge to filter irrelevant search results and web pages.

The paper is organized as follows: A summary of current medical information retrieval systems is given in Section 2. A background information on UMLS is given in Section 3. The structure of MedicoPort is presented in Section 4. Experimental results and discussions on MedicoPort are given in Section 5. Finally, possible future directions and conclusions are given in Section 6.

2. Background

Currently, there are four groups of gateways available to a medical information seeker: specialized guides in medicine, medical databases and search engines on these databases, general purpose search engines, and medical-domain web search engines.

Specialized guides in medicine are a structured compilation of medical resources maintained by health care experts. Since these guides are maintained manually by experts, the contents of the documents (web pages) are reliable, accurate and relevant to the topics. On the other hand, since the evaluation of documents is a slow process, the maintenance of these guides cannot cope with the growth rate of health information on the Internet [4]. An extension to medical guides is a subject directory where the users add documents to the hierarchy to achieve a more extensive coverage. As the health information grows, users fail to select the correct category to add and the search results become unpredictable. The web sites MedSurf, CliniWeb, and HardinMD shown in Table 1 are some examples of medical guides available on the Internet, and the web sites MedHelp, MedWeb, Medical Matrix are some examples for subject directories. These techniques, in general, lack of automation in maintenance to cope with the rapid growth of medical information on the web.

Table 1 – Medical information retrieval tools

Tool	Category	URL
MedSurf	Specialized guide	http://www.medsurf.com
CliniWeb	Specialized guide	http://www.ohsu.edu/clinweb
HardinMD	Specialized guide	http://www.lib.uiowa.edu/hardin/md
MedHelp	Subject directory	http://www.medhelp.org
MedWeb	Subject directory	http://www.medweb.emory.edu/MedWeb
Medical Matrix	Subject directory	http://www.medmatrix.org
MDConsult	Medical resource collection	http://www.medconsult.com
Medscape	Medical resource collection	http://www.medscape.com
PubMed	Medical database search	http://www.pubmedcentral.nih.gov/
MedTexus	Medical database search	http://ai.bpa.arizona.edu/go/medical/MedTexus.html
HelpfulMed	Medical Web search engine	http://ai.bpa.arizona.edu/helpfulmed
MedHunt/Marvin	Medical Web search engine	http://www.hon.ch/Medhunt
WRAPIN	Search on trusted sites	http://www.hon.ch/Project/Hunt_medical.html http://www.wrapin.org/
CMedPort	Medical Web search engine	http://ai.eller.arizona.edu/research/multilingual/cmepport.htm
Mamahealth	Medical web search engine	http://www.mamahealth.com/
Kosmix	Medical web search engine	http://www.kosmix.com
Healthline	Medical web search engine	http://healthline.com
HIQuA	Medical query assistant	http://samwise.bwh.harvard.edu/hiqua/hiqua.html

Download English Version:

<https://daneshyari.com/en/article/469791>

Download Persian Version:

<https://daneshyari.com/article/469791>

[Daneshyari.com](https://daneshyari.com)