Visualized Related Topics (VRT) System for Health Information Retrieval

Sukjin You
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
yous@uwm.edu

Joel DesArmo
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
jdesarmo@uwm.edu

Xiangming Mu
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
mux@uwm.edu

Sukwon Lee
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
sukwon@uwm.edu

Jessica C. Neal
University of Wisconsin-Milwaukee
P.O. Box 413,
Milwaukee, WI 53211
nealj@uwm.edu

ABSTRACT

To help bridge the gap between consumer user's vocabulary and controlled vocabulary used to index health information, in this demo we implemented a Visualized Related Topics (VRT) browser system. The VRT was integrated into the "MeshMed" [2] system to support health information retrieval. The key technology behind the VRT browser is to select MeSH terms, which represent the related topics or subjects, from the top relevant documents. We rank these MeSH terms using the traditional Term Frequency-Inverse Document Frequency (TF-IDF) algorithm. The VRT browser displays a graphic representation of these MeSH terms by creating a visual where the selected MeSH terms stem from the centered user query. The design goal is provide users an overview of the key topics of the search results. In addition, VRT browser may also help users form better queries. Using the VRT browser we will be studying how to effectively assist in consumer users with their health information seeking.

Keywords

Visualized related topics, health information retrieval system, related terms browser, related terms

1. INTRODUCTION

Consumer users usually experience difficulty when they use health information retrieval systems due to the vocabulary gap between the user's own query terms and the corresponding controlled vocabulary terms used to index the health information retrieval system [3]. The most common controlled vocabulary in health information retrieval is the Medical Subject Heading (MeSH).

Pseudo user feedback model and the thesaurus-based query expansion are commonly employed in information retrieval systems to assist in bridging such vocabulary gap in health information retrieval [1]. The thesaurus-based query expansion applies a thesaurus to map controlled vocabularies to user query terms. For example, the query expansion technology in the PubMed system will automatically add related MeSH terms to user's query. Continuous updates to reflect newly added user query terms is one limitation for this approach. In addition, a common concern for both the feedback models and the thesaurus-based query expansion technology is that they exclude the opportunity for user to select expanded query terms.

The consideration about how to provide "recommended terms" is an important topic for helping users in their information retrieval. There are many libraries and information providers who provide help regarding "recommended terms" in query formulation using various kinds of technologies. By selecting 10 MeSH terms that are associated with the top 50 most related documents, our research is to study the helpfulness and usefulness of VRT browser in health information retrieval systems. The selection of the MeSH terms is based on the TF-IDF scores calculated on the fly. These terms were presented in a visualized form which we called VRT browser (see Figure 1).

2. DESIGN PRINCIPLE

The health information retrieval system, "MeshMed,"[2] includes help features to support user's health information seeking. The VRT browser was developed and added to MeshMed as a help component in this Demo.

2.1 Data Set

We used dataset from an OHSUMED test collection (available at http://ir.ohsu.edu/ohsumed/ohsumed.html) which was intended to promote research on medical information retrieval. The OHSUMED test collection contains 348,566 references from MEDLINE which is a free available online database provided by National Library of Medicine. The collection includes several fields such as title, abstract, MeSH terms, author, source, and type of publication.

2.2 System Environment

We implemented the retrieval system using "Indri" engine that is a search engine developed by the University of Massachusetts and Carnegie Mellon University (http://www.lemurproject.org/indri/). To visualize the related terms, we used D3.js that is a JavaScript library developed by Mike Bostock (http://bost.ocks.org/mike/).

978-1-4799-5569-5/14/$31.00 ©2014 IEEE

2.3 Interface Design

We developed a new health information retrieval system with three components (see Fig. 2): (1) Search Browser with a query window like Google, (2) Tree Browser providing organized MeSH categorizations in a hierarchal structure, and (3) VRT browser presenting related MeSH terms around the original query term visually linked with lines. The length of each line is defined by the score of relevance ranking to the original query terms. The score is calculated using TF-IDF algorithm. For example, for the user's search query "Diabetes," some of the related MeSH terms are: "Diabetes Mellitus," "Diabetic Neuropathies," "Experimental," "Diabetic Nephropathies," "Diabetic Retinopathy," and "Diabetic Angiopathies." (See Figure 1)

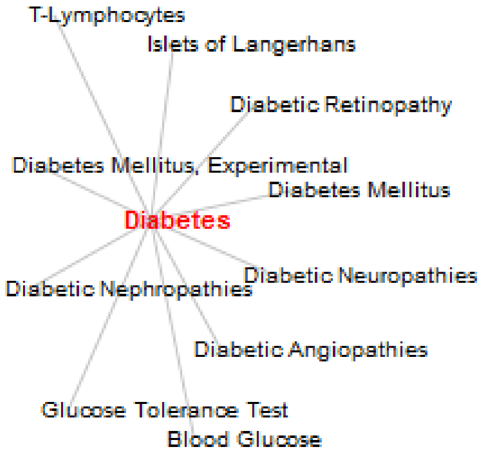


Figure 1. VRT browser.

These three components are inter-connected. For example, if a user clicks one of the recommended terms in the VRT browser, the tree browser will be updated to display the hierarchy of the selected term. If the user clicks one term in the tree browser, the search query term will be updated accordingly (see Figure 2).

3. CONCLUSION

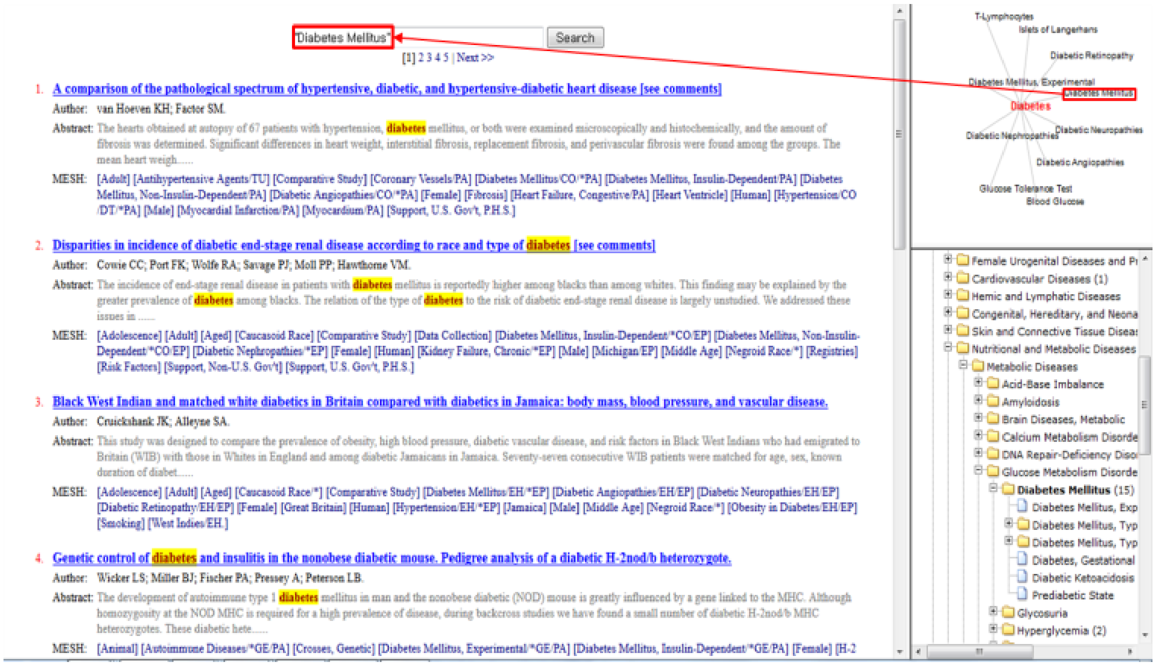
In health information retrieval, consumer users usually do not use the MeSH terms to form their queries. Recommending related MeSH terms to users might be helpful in finding more relevant documents. The Visualized Related Topics (VRT) browser was developed for the purpose of helping users' health information retrieval by providing MeSH terms that were selected from the top retrieved documents. The selection was based on the scores computed using TF-IDF algorithm. Our further study will explore the helpfulness and usefulness of the VRT in health information retrieval.

4. REFERENCES

[1] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*. Addison Wesley Longman, 1999, pp. 118-137.

[2] X. Mu et al., "Search strategies on a new health information retrieval system," *Online Information Review*, vol. 34, no. 3, pp. 440-456, 2010.

[3] Q. T. Zeng and T. Tse, "Exploring and developing consumer health vocabularies," *Journal of the American Medical Informatics Association*, vol. 13, no. 1, pp 24-29, 2006.



1. A comparison of the pathological spectrum of hypertensive, diabetic, and hypertensive-diabetic heart disease [see comments]
Author: van Hoeven KH; Factor SM.
Abstract: The hearts obtained at autopsy of 67 patients with hypertension, diabetes mellitus, or both were examined microscopically and histochemically, and the amount of fibrosis was determined. Significant differences in heart weight, interstitial fibrosis, replacement fibrosis, and perivascular fibrosis were found among the groups. The mean heart weigh......
MESH: [Adult] [Antihypertensive Agents/TU] [Comparative Study] [Coronary Vessels/PA] [Diabetes Mellitus/CO/*PA] [Diabetes Mellitus, Insulin-Dependent/PA] [Diabetes Mellitus, Non-Insulin-Dependent/PA] [Diabetic Angiopathies/CO/*PA] [Female] [Fibrosis] [Heart Failure, Congestive/PA] [Heart Ventricle] [Human] [Hypertension/CO/DT/*PA] [Male] [Myocardial Infarction/PA] [Myocardium/PA] [Support, U.S. Gov't, P.H.S.]

2. Disparities in incidence of diabetic end-stage renal disease according to race and type of diabetes [see comments]
Author: Cowie CC; Port FK; Wolfe RA; Savage PJ; Moll PP; Hawthorne VM.
Abstract: The incidence of end-stage renal disease in patients with diabetes mellitus is reportedly higher among blacks than among whites. This finding may be explained by the greater prevalence of diabetes among blacks. The relation of the type of diabetes to the risk of diabetic end-stage renal disease is largely unstudied. We addressed these issues in ......
MESH: [Adolescence] [Adult] [Aged] [Caucasoid Race] [Comparative Study] [Data Collection] [Diabetes Mellitus, Insulin-Dependent/*CO/EP] [Diabetes Mellitus, Non-Insulin-Dependent/*CO/EP] [Diabetic Nephropathies/*EP] [Female] [Human] [Kidney Failure, Chronic/*EP] [Male] [Michigan/EP] [Middle Age] [Negroid Race/*] [Registries] [Risk Factors] [Support, Non-U.S. Gov't] [Support, U.S. Gov't, P.H.S.]

3. Black West Indian and matched white diabetics in Britain compared with diabetics in Jamaica: body mass, blood pressure, and vascular disease.
Author: Cruickshank JK; Alleyne SA.
Abstract: This study was designed to compare the prevalence of obesity, high blood pressure, diabetic vascular disease, and risk factors in Black West Indians who had emigrated to Britain (WIB) with those in Whites in England and among diabetic Jamaicans in Jamaica. Seventy-seven consecutive WIB patients were matched for age, sex, known duration of diabet......
MESH: [Adolescence] [Adult] [Aged] [Caucasoid Race/*] [Comparative Study] [Diabetes Mellitus/EH/*EP] [Diabetic Angiopathies/EH/EP] [Diabetic Neuropathies/EH/EP] [Diabetic Retinopathy/EH/EP] [Female] [Great Britain] [Human] [Hypertension/EH/*EP] [Jamaica] [Male] [Middle Age] [Negroid Race/*] [Obesity in Diabetes/EH/EP] [Smoking] [West Indies/EH.]

4. Genetic control of diabetes and insulitis in the nonobese diabetic mouse. Pedigree analysis of a diabetic H-2nod/b heterozygote.
Author: Wicker LS; Miller BJ; Fischer PA; Pressey A; Peterson LB.
Abstract: The development of autoimmune type 1 diabetes mellitus in man and the nonobese diabetic (NOD) mouse is greatly influenced by a gene linked to the MHC. Although homozygosity at the NOD MHC is required for a high prevalence of disease, during backcross studies we have found a small number of diabetic H-2nod/b MHC heterozygotes. These diabetic hete......
MESH: [Animal] [Autoimmune Diseases/*GE/PA] [Crosses, Genetic] [Diabetes Mellitus, Experimental/*GE/PA] [Diabetes Mellitus, Insulin-Dependent/*GE/PA] [Female] [H-2

Figure 2. Search Browser (left) + VRT browser (top right) + Tree browser (bottom right).