ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

Developing Trends in Bio-medical Search Engines

Amita Murthy[1], Dr. Padmaja K. V.[2]

Student, Instrumentation Technology, R. V. C. E., Bengaluru, India[1]

Professor and Associate Dean, Instrumentation Technology, R. V. C. E., Bengaluru, India[2]

Abstract: Complete access to the existing pool of biomedical literature and the ability to "hit" upon the exact information of the relevant specialty are becoming essential elements of academic and clinical expertise. With the rapid expansion of the literature database, it is almost impossible to keep up to date with every innovation. Hence a review about the state-of-art technologies in bio-medical search engines has been presented. The paper provides a brief summary about the developments so far in factors like technologies used, methodologies implemented, and the new features these search engines provide. The paper also gives us a clear picture about the commercially available products and the possible future scope in this area.

Keywords: Oncology graph, Semantic similarity, Google AJAX, Predicate based Vector Space Model

I. INTRODUCTION

In the present era of information and technology, staying up to date with the latest advances in biomedical sciences is a major challenge for clinical practitioners. Because the amount of biomedical information doubles every five years, clinicians must have free and easy access to the current literature database for easy and effective evidence based clinical decision-making [1]. Traditionally, there have been several systems available that condense and dispense the medical intelligence in to easily absorbable forms (e.g. medical and dental textbooks and dictionaries). However, these are frequently based on the synopsis and ideas of established experts and may not be refreshed with current information. In our day-to-day practice, we often come across a single and specific clinical problem that may be explained well in a single article. Until recently, the problem for many clinicians has been accessing this information.

The World Wide Web or Internet has resolved these dilemmas to a large extent. Its rapid growth has created a boom in the field of biomedical investigation and research, although there is a long way to go before its full potential is realized [2,3]. On the click of mouse/button, the Internet offers quick and economical access to medical literature in the form of databases, dictionaries, journals, textbooks, dental product information, continuing education resources, and e-journals [4,5]. There are several eminent interlinking sources that retrieve these results such as government organizations, publishers, online health libraries, and commercial agencies. Searching biomedical literature is a very organized and specific procedure. It requires systematic planning so as to develop a well-constructed clinical question or precise keyword. Unplanned and messy efforts may result in the retrieval of several, apparently irrelevant articles thus discouraging the professional to look further [6-8]. Web based search engines are tools designed specially to search for information in the form of images, databases, journals and dictionaries. As these search engines are computer operated, they mostly search algorithmically. With these points in mind, it is therefore very important to understand how to access the information that is being searched for. This paper provides a detail description of the inventions done in bio-medical search engine field. The paper is divided into three sections. The first section lists out the important units in bio-medical search engines and the latest inventions. The second section gives us the possible future scope in this area. The third section concludes the discussion in this paper.

II. UPCOMING TECHNOLOGIES

Yuanyuan Zhang et al.[9] developed a G-Bean: an Ontology-graph based Web Tool for Biomedical Literature Retrieval. The aim of the project was to search for biomedical documents from the MEDLINE database. This search engine used an ontology-graph based query expansion technique to expand the initial query with additional query terms and/or more specific query concepts, resulting in the retrieval of more accurate and relevant information from the database. A multithreading parallel process is used to automatically generate the document index to address the inefficiency of the PubMed's manual indexing process. This search engine could also discover user's true search intention and retrieve additional relevant articles based on articles that he/she has already shown interest in. The methodology was: Index creation, Query Expansion, Document Retrieval. Fig. 1 shows the block diagram of the system

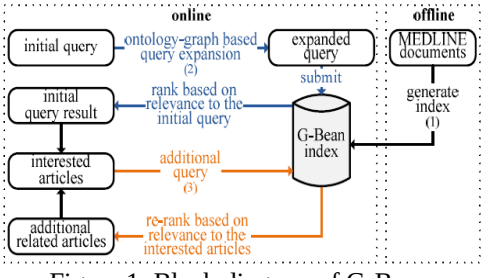

Figure 1: Block diagram of G-Bean

ISSN (Online) : 2278-1021
ISSN (Print) : 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 3, Issue 5, May 2014

Chi-Huang Chen et al.[10] developed Semantic Similarity Measure in Biomedical Domain Leverage Web Search Engine. Semantic similarity measure plays an essential role in Information Retrieval and Natural Language Processing. In this paper page-count-based semantic similarity measure has been proposed and applied it in biomedical domains. Previous researches in semantic web related applications have deployed various semantic similarity measures. Despite the usefulness of the measurements in those applications, measuring semantic similarity between two terms remains a challenge task. The proposed method exploits page counts returned by the Web Search Engine. We define various similarity scores for two given terms P and Q, using the page counts for querying P, Q and P AND Q. These different similarity scores are integrated adapting support vector machines, to leverage the robustness of semantic similarity measures. Experimental results on two datasets achieve correlation coefficients of 0.798 on the dataset provided by A. Hliaoutakis, 0.705 on the dataset provide by T. Pedersen with physician scores and 0.496 on the dataset provided by T. Pedersen et al. with expert scores.

Sheau-Ling Hsieh et al.[11] developed Semantic Similarity Measures in the Biomedical Domain by Leveraging a Web Search Engine. The traditional ontology-based methodologies have a limitation that both concepts must be resided in the same ontology tree(s). Unfortunately, in practice, the assumption is not always applicable. On the other hand, if the corpus is sufficiently adequate, the corpus based methodologies can overcome the limitation. Now, the web is a continuous and enormous growth corpus. Therefore, a method of estimating semantic similarity is proposed via exploiting the page counts of two biomedical concepts returned by Google AJAX web search engine. The features are extracted as the co-occurrence patterns of two given terms P and Q, by querying P, Q, as well as P AND Q, and the web search hit counts of the defined lexico syntactic patterns. These similarity scores of different patterns are evaluated, by adapting support vector machines for classification, to leverage the robustness of semantic similarity measures. Experimental results validating against two datasets: dataset 1 provided by A. Hliaoutakis; dataset 2 provided by T. Pedersen. In dataset 1, the proposed approach achieves the best correlation coefficient (0.802) under SNOMED-CT. In dataset 2, the proposed method obtains the best correlation coefficient (SNOMED-CT: 0.705; MeSH: 0.723) with physician scores comparing with measures of other methods. However, the correlation coefficients (SNOMED-CT: 0.496; MeSH: 0.539) with coder scores received opposite outcomes.

Myungjae Kwak et al.[12] performed a Pilot Study of a Predicate-based Vector Space Model for a Biomedical Search Engine. A search engine that supports finding precise biomedical statements, but also complementary, and contrasting information would greatly help biomedical researchers. The proposed system uses predicates in a search engine's underlying data structure to accomplish this. A predicate is a triple that combines two phrases with a predicate. The evaluation of the search engine was conducted by comparing three different approaches: keyword-based search, triple-based search, and an additive search that combines keywords and predicates. Cancer researchers provided the queries, relevant to their ongoing work, and evaluated the outcome in a double-blind fashion. The results showed that the combined approach, which combines triple-based and keyword-based approaches, always outperformed the 2 other approaches.

Michael Gubanov et al.[13] developed READFAST: Optimizing Structural Search Relevance for Big Biomedical Text. The aim of this project is to find relevant data from large textual information, like Books; Journals; Electronic Health Records; etc. The methodology involves three major steps i.e. creation of database with biomedical books, journals and records and other big biomedical data. Then the major sections of the documents are identified and stored in the creation of structured index step. The sections are further scanned and sub-sections are created based on the importance of the data in it, this is done in the creation of attributes step. The end user first searches for the index terms and then the attributes. The success of the search engine is measured using Normalized Discounted Cumulative Gain (NDCG). Average NDCG gain for all queries for the first test corpus was 23.2%, 30% for the second corpus, which is quite a significant value. Fig. 2 shows the block diagram of the system

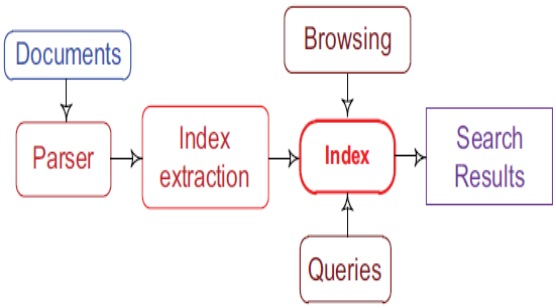

Figure 2: Block diagram of READFAST

III. COMMERTIALLY AVAILABLE PRODUCTS

Many leading companies across the world have been developing various types of search engines for searching the biomedical literature. 'PubMed' is the US Government supported effort to provide a means to search the biomedical literature. Over the years, this amazing effort has been critical for essentially every research project and for most physicians[14]. 'Google Scholar' is a cross-disciplinary search engine for journal articles and other scholarly works[15]. 'Quertle' is a rapidly growing option in the biomedical search engine area. Quertle is the only major biomedical search engine to focus on semantic searching to improve the relevance of results. Like PubMed and Google Scholar, Quertle is freely accessible[16]. 'BioMedSearch.com' is a biomedical search engine that contains NIH/PubMed documents, a large collection of theses, dissertations, and other proprietary publications, making it one of the most comprehensive and powerful free biomedical searches [17].

IV. FUTURE SCOPE

In near future, the project can be extended to build hospital workflow demonstration tools. It has been found that about 60% doctors and nursing staff find it difficult to browse through the healthcare IT solutions as they have limited technical knowledge and the information is scattered across various resources and is not found at a single location. Hence development of a standalone biomedical search engine application which retrieves information about various medical scenarios in hospitals, the description and specifications about the scenarios and also their workflow would be of great help.

V. CONCLUSION

The purpose of this study was to provide a review about the existing technologies in the field of cardiac monitoring devices. Hence a literature survey has been done and a brief idea about the technologies developed in this domain has been projected. In addition to the existing technologies, a novel idea about the future scope in this area has also been provided.

REFERENCES

[1] Mattox DE. Welcome to ARCHIVES CME. Archives of Otolaryngology—Head and Neck Surgery. 2000; **126**: 914.
[2] Jenner A. How has the Internet affected dentistry? [Research summary of: Chestnutt IG, Reynolds K. Perceptions of how the Internet has impacted on dentistry]. British Dental Journal 2006; 200: 149.
[3] Schleyer TK, Spallek H, Torres-Urquidy MH. A profile of current Internet users in dentistry. Journal of the American Dental Association. 1998; **129**: 1748-1753.
[4] Doyle DJ, Ruskin KJ, Engel TP. The Internet and medicine: Past, present, and future. Yale Journal of Biology and Medicine. 1996; **69**: 429-437.
[5] Jethwani KS, Chandwani HS. The internet: Revolutionizing medical research for novices and virtuosos alike. Journal of Postgraduate Medicine. 2008; **54**: 49-51
[6] Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence- Based Medicine Working Group. Journal of the American Medical Association. 1993; **270**: 2093-2095.
[7] Bin Ghouth AS. Using computer and Internet for medical literature searching among medical students in Hadramout University, Yemen. Online Journal of Health and Allied Sciences. 2008; 7(1): 6. Available from: http://www.ojhas.org/ issue25/2008-1-6.htm
[8] Stewart MG, Moore AS. Searching the medical literature. Otolaryngologic Clinics of North America. 1998; **31**: 277-287.
[9] Yuanyuan Zhang, Liang Dong, Lin Li, Pradip K. Srimani, Philip S. Yu, James Z. Wang, "G-Bean: an Ontology-graph based Web Tool for Biomedical Literature Retrieval", 2013 IEEE International Journal on Bioinformatics and Biomedicine
[10] Chi-Huang Chen, Sheau-Ling Hsieh, Yung-Ching Weng, Wen-Yung Chang, Feipei Lai, "Semantic Similarity Measure in Biomedical Domain Leverage Web Search Engine", IEEE Journal Of Biomedical And Health Informatics, Vol. 32, No. 2, September 4, 2010
[11] Sheau-Ling Hsieh, Wen-Yung Chang, Chi-Huang Chen, Yung-Ching Weng, "Semantic Similarity Measures in the Biomedical Domain by Leveraging a Web Search Engine" , IEEE Journal Of Biomedical And Health Informatics, Vol. 17, No. 4, July 2013
[12] Myungjae Kwak, Gondy Leroy, Jesse D. Martinez, "A Pilot Study of a Predicate-based Vector Space Model for a Biomedical Search Engine", 2011 IEEE International Conference on Bioinformatics and Biomedicine Workshops.
[13] Michael Gubanov, Anna Pyayt, "READFAST: Optimizing Structural Search Relevance for Big Biomedical Text", IEEE IRI 2013, August 14-16, 2013, San Francisco, California, USA.
[14] US National Library of Medicine National Institutes of Health, "http://www.ncbi.nlm.nih.gov/pubmed", Copyright october 28, 2009.
[15] Google scholar, "http://scholar.google.co.in/"
[16] Quertle, "http://www.quertle.info/?cmp=lsse"
[17] A Sumobrain solution company, "http://www.biomedsearch.com/", Copyright 2009-2010