MLM-based Automated Query Generation for CDSS Evidence Support

Muhammad Afzal, Maqbool Hussain, Taqdir Ali, Wajahat Ali Khan, Sungyoung Lee, Byeong Ho Kang

{muhammad.afzal, maqbool.hussain, taqdir.ali, wajahat.alikhan, sylee}@oslab.khu.ac.kr,
Byeong.Kang@utas.edu.au

Abstract. Clinical decision support system (CDSS) is fast becoming a requirement in diverse medical domains to assist physicians in clinical decisions. Physicians look at the research evidences for satisfaction in CDSS assisted clinical decisions and also to keep their knowledge up-to-date. Research evidences are available in the form of studies, summaries, and other formats published in credible journals, books and reviews as online sources. The most important and critical part to get the evidences in a better way is the search query generation and its optimization. A query that is characterized by domain context and clinical workflow, and optimized for the target search engine in order to generate right and relevant results. In most cases, the search queries are generated manually, which require a lot of physicians' time to get the right information. Other follow automated way of generating queries from electronic medical records, which make it difficult to associate evidences to the clinical decisions. The role of the source from where the queries are created is highly important. We are presenting the work of query generation from Medical Logic Modules (MLMs) as a main source of query contents. We create different query set from the concepts used in MLMs expended with domain ontology derived from SNOMED CT. The results are compiled with respect to coverage using classified training set of over 380 research articles. The proposed work is demonstrated to physicians and their feedback upon time saving as well as presentation of information in the context was highly positive.

Keywords. CDSS, Query Generation, Query Expansion, Evidence Support

1 Introduction

Over the last few decades, the ever-increasing output of scientific publications has introduced new requirements for professionals, e.g., physicians, who have to locate the exact papers that they need for their clinical research work amongst a huge number of publications [1]. Only MEDLINE/PubMed Baseline yearly citations totals from 2014 are 22,376,811 reported on in statistical reports on MEDLINE/PubMed Baseline data by U.S. National Library of Medicine [2]. The research paper searched by the

physicians are mainly based on the input queries. The contextual enriched and pertinent queries results in better output as compared to ad hoc queries.

In most cases, the search queries are generated manually, which require a lot of physicians' time to get the right information. Other follow semi-automated and automated ways of generating queries from electronic medical records or electronic health records [5, 6, 7, 8], which make it difficult to associate evidences to the clinical decisions. The role of the source from where the queries are created is highly important. There is lack of associating clinical decision, made by clinical decision support system (CDSS) with research articles as evidence in order to get satisfaction over the system decisions and keep up-to-date with research literature. An approach is required that can automatically generate contextual queries able to find right and relevant evidences to support the clinical decisions made by CDSS.

We are presenting the work of query generation from Medical Logic Modules (MLMs) [3] as a main source of query contents. MLM is Arden Syntax encoded representation containing sufficient knowledge to make a single decision. The system generate the queries with two different strategies; MLM only queries, MLM-Domain Ontology combined queries. Within these two strategies, we identify several variations to formulate query and test results against each method. Each query returned results in variation and relevancy is checked with research articles in training set.

2 Methodology

This work is undertaken as a part of a Smart CDSS system [9]. Smart CDSS consists of three major components with several subcomponents. The three components are; knowledge authoring, knowledge base, and research evidence support. ***Knowledge authoring component*** provides the environment for physicians to create their clinical knowledge in the form of knowledge rules. The knowledge rules are created using domain ontology derived from SNOMED CT. It has the verification and compilation subcomponents for verifying and compiling the rules [10]. ***Knowledge base*** maintains the knowledge rules in the form of Medical Logic Module (MLM). MLM encapsulates knowledge as software module that triggers an action based on data event generated at healthcare system [3]. Initially the MLM was intended to have single logic that act on single set of data and result in single set of actions. However, now it can support to invoke other MLMs that result in chaining of actions with its own logic and set of data elements. Each MLM contains slots which are logically grouped into three required categories and one optional category; maintenance, library, knowledge and resources (optional).

For query generation, we are utilizing the "knowledge" category of MLM that contains slots to specify intention of what MLM does. Its sub-slot includes data slot (define terms used in MLM), evoke slot (specify context of MLM evocation), logic slot (the actual condition to be tested on terms) and the action slot (specify the action that should be taken in case condition is true). The third part i.e. ***research evidence support*** provides the mechanism to incorporate relevant articles as evidences for the decisions of CDSS [11, 12]. There are two models of this mechanism; push model and

pull model. In push model, the evidence incorporation is automatic without any involvement of human while in pull model, semi-automatic approach is applied with human participation. In this paper, the focus is push model where we are generating the queries from MLMs of knowledge base components automatically. The process of generating query automatically to search evidences from online sources involves three important steps as shown in Fig. 1.

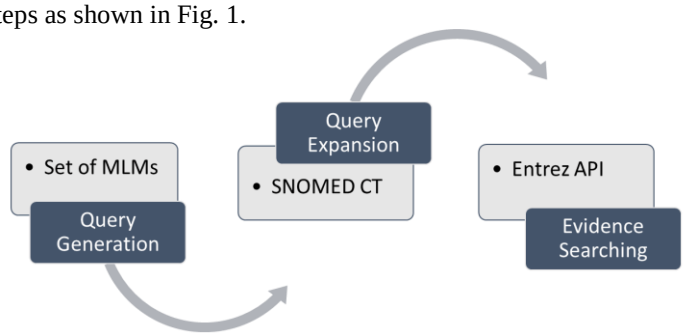

Fig. 1. Query generation, expansion, and evidence searching model

Query Generation utilizes the actuated MLMs that participated in clinical decision and build the query from the terms embodied in logic part of the MLMs. ***Query Expansion*** expands the terms with standard vocabulary of SNOMED CT for the purpose to increase the relevant results coverage. ***Evidence Searching*** utilizes Entrez Programming Utilities (eUtils), a stable interface into the Entrez query and database system at the National Center for Biotechnology Information (NCBI) used by PubMed search engine.

The system has been developed for head and neck cancer domain and tested different patient cases by generating both simple and expanded queries. The average recall score of simple queries was 75% while for expanded queries it was 95%. It proves that expansion of query terms plays an important role to consider for better results.

3 CONCLUSION

Evidence support is among the important parts of any contemporary CDSS system. It not only increases the satisfaction level of users including physicians, nurses and clinical researchers rather also provide a coherent way to keep the knowledge base up-to-date. In the form of MEDLINE we have access to more than 21 million reviewed research articles in medical domain. The most critical obstacle is the coherent integration of research evidence with the decisions of CDSS system. We presented query generation model containing three primary steps: query generation, query expansion and evidence searching. The results are evaluated with the help of training set having classification of 380 research articles for different anatomical sites of head and neck cancer. The results proved that expanded query provides better coverage as compared to simple non-expanded query. The automation of evidence searching has reduced much of the physicians' time spent on designing manual queries.

Acknowledgement

This work was supported by the Industrial Strategic Technology Development Program (10035348, Development of a Cognitive Planning and Learning Model for Mobile Platforms) funded by the Ministry of Knowledge Economy(MKE, Korea)

REFERENCES

[1] Perez-Rey, David, et al. "CDAPubMed: a browser extension to retrieve EHR-based biomedical literature." BMC medical informatics and decision making 12.1 (2012): 29.

[2] Statistical Reports on MEDLINE/PubMed Baseline Data http://www.nlm.nih.gov/bsd/licensee/baselinestats.html [Accessed on 25th Jan, 2014]

[3] Ohno-Machado L, et al. The GuideLine Interchange Format A Model for Representing Guidelines. *Journal of the American Medical Informatics Association*. 1998; 5: 357-372.

[4] Sayers, Eric, and David Wheeler. "Building Customized Data Pipelines Using the Entrez Programming Utilities (eUtils)." (2004).

[5] Cimino, James J. "An integrated approach to computer-based decision support at the point of care." Transactions of the American Clinical and Climatological Association 118 (2007): 273.

[6] Perez-Rey, David, et al. "CDAPubMed: a browser extension to retrieve EHR-based biomedical literature." BMC medical informatics and decision making12.1 (2012): 29.

[7] Chuang, Shui-Lung, and Lee-Feng Chien. "Automatic query taxonomy generation for information retrieval applications." Online Information Review 27.4 (2003): 243-255.

[8] Kim, Yang Sok, and Byeong Ho Kang. "Search query generation with mcrdr document classification knowledge." Knowledge Engineering: Practice and Patterns. Springer Berlin Heidelberg, 2008. 292-301

[9] Maqbool Hussain, Wajahat Ali Khan, Muhammad Afzal, Sungyoung Lee, "Smart CDSS for Smart Homes", 10th International Conference On Smart homes and health Telematics(ICOST 2012), Artimino, Italy, June 12-15, 2012.

[10] Taqdir Ali, Maqbool Hussain, Wajahat Ali Khan, Muhammad Afzal and Sungyoung Lee, "Authoring Tool: Acquiring sharable knowledge for Smart CDSS", 35th IEEE Annual International Conference of the Engineering in Medicine and Biology (EMBC'13), July 3-7, 2013.

[11] Afzal, Muhammad, et al. "Meaningful Integration of Online Knowledge Resources with Clinical Decision Support System." Inclusive Society: Health and Wellbeing in the Community, and Care at Home. Springer Berlin Heidelberg, 2013. 280-285.

[12] Muhammad Afzal, Maqbool Hussain, Wajahat Ali Khan,Taqdir Ali, Sungyoung Lee and Byeong Ho Kang, "KnowledgeButton: An Evidence Adaptive Tool for CDSS and Clinical Research", International Symposium on INnovations in Intelligent SysTems and Applications (INISTA'14), June 23-25, 2014