International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 03 Issue: 07 | July-2016 www.irjet.net p-ISSN: 2395-0072

Eliminating the Gap between Health Conscious people and Health Care based Knowledge Sharing Portal

Sonali Porwal[1], Prof. Sunanda Biradar[2]

[1]M.Tech Student, *Department of Computer Science and Engineering*
BLDEA'S V.P. Dr. P.G. Halakatti College of Engineering and Technology, Vijayapur
[2] *Assistant Professor, Department of Computer Science and Engineering*
BLDEA'S V.P. Dr. P.G. Halakatti College of Engineering and Technology, Vijayapur

---

Abstract - *Web usage is increasing day by day for the purpose of health seeking and health care. There is also a tremendous increase in the number of medical records available on the web over the years. There may be vocabulary difference between the user who posts the query and the content available on the web. Due to this difference in vocabulary, it becomes difficult to give accurate results for the searched query from such a large collection of medical records. Therefore, this system automatically maps the terminology used in the query to corpus aware terminology. The system uses two approaches: the local mining and the global mining. In the local mining phase, the query is optimized locally, in which the medical terms are extracted and mapped to corpus aware terminology. But, local mining results are not accurate and suffer from loss of information, therefore global mining is used. Global mining enhances the local mining results. Global mining has two components: inter-expert relationship and inter-terminology relationship. The system has the ability to handle large volumes of data.*

***Key Words*: local mining, global mining, vocabulary gap etc.**

1.INTRODUCTION

With the advancement in the internet technologies, the era for medical treatment is also changing day by day. Healthcare seekers and healthcare providers are becoming dependent on the online health seeking and looking for various medical advices, treatments, suggestions and education from the internet. Patients are actively seeking for online health care tips. With the various surveys conducted worldwide, it was found out that the percentage of people seeking for heath concerns is increasing day by day.

With such a growing trend for online health seeking, many online web portals have been launched where the patient or a health seeker could interact worldwide with the specialized doctors through Question Answers (QA). Also frequently asked questions are mentioned on such portals from where general queries could be easily solved. This trend is beneficial to both the professionals as well as the patients or health seekers. The doctors get a platform where they get a chance to display their knowledge which in turn increases their reputation and they can also interact with other renowned doctors. Patients, on the other hand, get a platform where they can easily ask for their queries and get authenticated answers. This is a very easy way for interaction with the specialist. Web contains a large set of medical reports from where patients can directly get their answers and they do not have to wait for the expert to reply.

Many a times, the web generated content is not directly usable as the query that the patient posts does not contain the same vocabulary as in the medical corpus. Thus, due to the vocabulary gap such large corpus of authenticated medical information is unused and the patients do not get satisfactory results. Also, the professionals use standardized medical terms which the patient is not aware off. As a consequence, the search results are not appropriate and satisfactory. Consider HealthTap as an example which is a health related site where the patients can ask question of their concern. The questions are not in the standardized terms and two different patients can ask the same questions in different terminologies. Also, the doctors answer the questions with non standardized terms and may contain words that have multiple possible meanings. Data exchange, data usage, integrity and management become difficult due to the presence of inconsistent community generated data. The users also found it difficult and challenging to use the inconsistent data as there was a difference in the terms of the query they post and those medical records that are stored in the corpus. All these reasons motivate us for building new software to eliminate this gap. The medical records should be coded with standardized technologies so that it becomes easy to index, store and aggregate across the sites.

The proposed system automatically codes medical records with the help of corpus aware terminologies. The system uses two types of mining: Local mining and Global Mining. Local mining codes the queries submitted by the user locally by extracting local information from the query itself. The information that is extracted from the query is

the medical concept in the query. These are mapped to the corpus aware terminologies. Local mining consists of three phases which are: Noun Phrase Extraction, Medical Concept Detection, Medical Concept Normalization. In the noun phrase extraction phase the nouns are extracted from the query. In Medical Concept Detection, the medical concepts are extracted from the nouns extracted in the previous phase. In Medical Concept Normalization, the medical terms are mapped to the corpus aware terminologies. The result of local mining acts as the input to the second phase that is the global mining. Global mining further enhances the results of the local mining as in local mining the results are less accurate. Local mining results suffer from loss of information and low precision, therefore for more accurate results, the local mining results are then passed for global mining. Global mining extracts missing concepts. The components of global mining are: the inter-terminology relationships and the inter-expert relationships. The inter-terminology relationships make use of well structured dictionary to extract the useful information. The inter-expert relationship makes use of the expert's historical data.

2. RELATED WORK

The current medical records are coded and organized manually [1]. This is a very tedious task and requires lots of efforts. Therefore, automatic system needs to be developed to convert the terminology to corpus aware terminology. There are two approaches for automatic assignment of medical terms: Rule based approach and Machine Learning approach. Health seekers these days seek for their better health online. Generally, when a health seeker searches through some query, matched Question-Answers are returned. This may not satisfy the seeker. Therefore, the authors introduced a system which supports multiple languages and obtains a well structured result, which is multi-faceted, from many heterogeneous sources. If the health seeker is not satisfied with the result, the system sends the query of the user to the professionals with appropriate expertise automatically. Wenzer represents the results obtained for the query into a single view automatically. The system provides three different views which are patient generated data, expert generated data and the health provider released data. [2] International classification of diseases is the alphanumeric codes assigned for the diagnosis of various kinds of diseases which is very useful for reimbursement of the bills. Assigning codes to the discharge summaries manually is quite hectic and requires great deal of accuracy. So automatic assignment should be implemented. This kind of automatic classification is done by using three kinds of classifiers which are: Bayesian independence classifiers, k-nearest neighbor classifiers and relevance feedback. All of these classification approaches are Rule based approaches. In the k-nearest neighbor approach, an already trained coded document is used. The document to which the code is to be assigned is used as the test document. Many experiments were performed to maximize the output for the desired test document that is assigning correct cod to the discharge summaries. These classification algorithms use the various concepts of natural language processing to form the various rules for classification like semantic, syntactic, morphological and pragmatic aspects. These rule based approaches are used widely these days for automatic assignment of medical terminologies. [3] The authors in [4], proposed a tool called as medical concept mapper, for easy access to online medical information. The users post a medical related query and the tool then produces appropriate related medical terms for their query that can further be used for the purpose of searching for their medical related problem. The tool provides great help to the people who have less information of the appropriate medical terms to be used while searching for their medical problem or for experts who have to search a medical query but beyond their expertise. The mapper produces the appropriate medical terms by producing synonyms of the words in the query and by the terms semantically to the related concepts in the query. The system is different in the way it combines its components which are the Arizona noun phraser (AZ), which contains ontologies created by human beings, Unified Medical Language System and WordNet, and also the concept space created into the computer. The authors did a comparative study and found out that the Arizona Noun Phrase extractor is a good choice for refining the medical terms from the user query. WordNet does not provide satisfactory results as the metathesaurus provides in finding out the synonyms of the medical terms or phrases extracted. [4] Both health care practitioners and academic researchers use the applications of natural language processing methods for clinical text free data. These applications of natural language processing are used to provide a strong support in the areas of science, education, decision making and administration. The authors have introduced a system which was capable of automatically assigning diagnostic codes to the radiology reports using multi-label classification. The system for such automatic assignment consisted of feature selection, text enrichment and two classifiers. The text enrichment and the feature selection method increased the accuracy of the system a lot and helped in achieving the desired target of the system. The system uses the concepts of machine learning. Amongst all of the components of the system, feature engineering components provided a challenging edge at the international level. Also amongst the two classifiers used, if the first classifier made some mistake that is known in advance, then the output of the second classifier is used to give the result of the final prediction. [5]

3. METHODOLOGY

The proposed method aims to minimize the vocabulary gap between the medical terms used by the user in the query and the corpus aware terminology. It maps the medical terms used in the query to the corpus aware terminology.

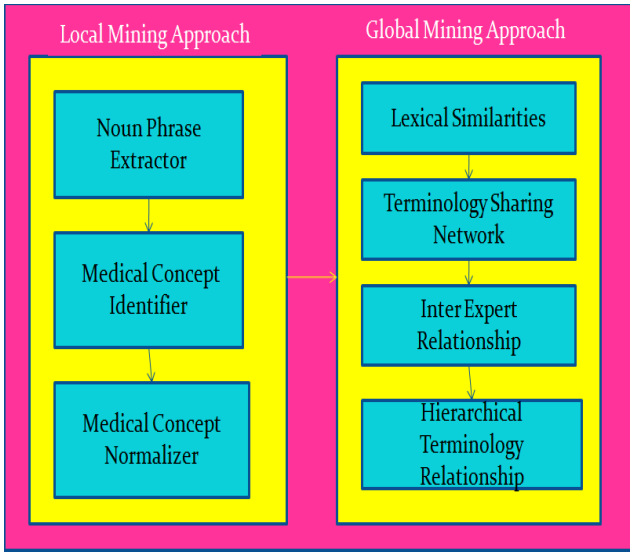

Fig. 1 System diagram

The system uses two kinds of mining which are: Local Mining and the Global Mining. Local mining deals with the query that the user submitted to the system. The useful medical information that is the medical terms needs to be extracted from the submitted query. Local mining forms a tristage framework that includes three steps which are: Noun Phrase Extraction, Medical Concept identifier and Medical Concept Normalizer. Firstly, the user submits the query to the system. In the noun phrase extraction phase, the noun phrases contained in the query is extracted using NLP and all the other non medical terms are removed like stopwords, etc. Then, the medical terminologies are extracted from the noun phrases that are extracted from the previous phase and other non medical terms are ignored. After that normalization of the medical terms is performed. The results of the local mining are passed as the input for global mining. Therefore, local mining should be performed in such a way that the result of it enhances our global mining results. Global mining has two phases which are inter-terminology relationship and inter-expert relationship. A hierarchical structure is used to organize the medical terminologies in SNOMED CT. For example "viral pneumonia" is generalized as "infectious pneumonia" which is further generalized as "pneumonia". Terminologies can have many parents. If the experts that we consult are professionals in that field or related fields, then the results obtained would be more beneficial. In that case, the inter-expert relationship is viewed as stronger

4. RESULTS

The system is designed in way that its efficiency is maximum when both local and global mining approaches are used together. When local mining approach is used alone, some works from the experts are missed out. The efficiency of the system is less. If global mining is used after the local mining approach, the efficiency of the system is increased.

The efficiency of the system can be calculated as:

System efficiency: (No. of queries correctly classified/Total no. of queries tested) *100

The local mining approach optimizes the query locally and global mining enhances the results of the local mining approach. The final result of system is the expert advice for the query submitted by the user and the corpus aware terminology for the medical terms used by the user. The results also shows the hierarchical terms for the medical terms contained in the query.

The efficiency of the system by using only local mining approach is low and therefore global mining approach is collaborated in the system. The system has the capability to hold large volumes of data.

The efficiency of system is calculated as follows:

System efficiency= (15/20) *100

=75%

Therefore, conclusion can be made that using both local and global mining approach maximizes the efficiency of the system.

5. CONCLUSIONS

The proposed system automatically codes the query submitted by the user to the corpus aware terminologies. The system consists of two phases: local mining phase and the global mining phase. The first phase, local mining evaluates the query locally and extracts useful medical terminologies from the query that the user of the system posts. This phase consists of three different components for coding the query locally: noun phrase extraction, medical concept identifier and medical concept normalizer. This phase may suffer from less accurate information and low precision. Therefore, the second phase was introduced, the global Mining, which enhances the results of local mining phase. This phase checks for

relationship management like inter-expert relationship and inter-terminology relationship. This whole system is not under any supervision and it has the capability to hold large amounts of data. The proposed system can be used effectively by general people in their day to day life for searching for specific health related problems.

REFERENCES

[1] AHIMA e-HIM Work Group on Computer-Assisted Coding, "Delving into computer-assisted coding," J. AHIMA, vol. 75, pp. 48A–48H, 2004.

[2] L. Nie, T. Li, M. Akbari, and T.-S. Chua, "Wenzher: Comprehensive vertical search for healthcare domain," in Proc. Int. ACM SIGIR Conf., 2014, pp. 1245–1246.

[3] L. S. Larkey and W. B. Croft, "Automatic assignment of icd9 codes to discharge summaries," PhD dissertation, Dept. Comput. Sci., Univ. Massachusetts at Amherst, Amherst, MA, USA, 1995.

[4] G. Leroy and H. Chen, "Meeting medical terminology needs-the ontology-enhanced medical concept mapper," IEEE Trans. Inf. Technol. Biomed., vol. 5, no. 4, pp. 261–270, Dec. 2001.

[5] H. Suominen, F. Ginter, S. Pyysalo, A. Airola, T. Pahikkala, S. Salanter, and T. Salakoski, "Machine learning to automate the assignment of diagnosis codes to free-text radiology reports: A method description," in Proc. ICML Workshop Mach. Learn. Health-Care Appl., 2008