Sanative Chatbot For Health Seekers

V.Manoj Kumar[1], A.Keerthana[2], M.Madhumitha[3], S.Valliammai[4], V.Vinithasri[5]

Department Of Information Technology,Saranathan College Of Engineering, Trichy,Tamilnadu,India.

A*bstract*: **Now a day people tend to seek knowledge or information from internet that concern with health through online healthcare services. The basic aim of this system is to bridge the vocabulary gap between the health providers by proving instant replies to the questions posted by patients. Automatic generated content for healthcare services are chosen instead of traditional community generated systems because they are reliable, compatible, and provide instant replies. This paper proposes a scheme to code the medical record using local mining and global approaches. Local mining aims to code the medical records by extracting the medical concepts from individual record and then mapping them to terminologies based on external authenticated vocabularies. Local Mining establishes a tri-stage framework to accomplish this task. Global learning aims to learn missing key concepts and propagates precise terminologies among underlying connected records over a large collection.**

Index Terms - Local mining, global learning, corpus aware terminology, POS tagger

I. INTRODUCTION

Recent years people get addicted to internet in obtaining information for every problem they face. This not only yet people to seek knowledge about general topics but also their personal health concerns. This not only let people to seek knowledge about general topics but also their personal health concerns. This was confirmed by a national survey conducted by Pew Research Centre in Jan 2013, where they reported that one in three American adults have gone online to figure out their medical conditions in the past 12 months from the report time. This also shows patients who struggle usually explore the internet to research the problem before and after they see their doctors.

Establishing question answer forums is becoming a simple way to answer those queries rather than browsing through the list of potentially relevant document from the web. The forums left many questions unanswered in the portal. The need for reliable communication and to better cater the health seekers from getting good solutions, a number of community-based healthcare services have turned up. The importance liein answering the questions of patients at right time. For this reason online services emerged with better efficiency.

II. EXISTING SYSTEM

In many cases, the solutions provided by doctors would not be usable to the patients due to inclusion of medical concepts and terminologies. The answers provided by the well trained experts may contain acronyms with multiple possible meanings and non-standardized terms. The users struggle to understand those terminologies and have a overhead. The inconsistency of community generated health content greatly hindered data exchange, management and integrity. The users have faced big challenges due to incompatibility in vocabulary being used in question answer pairs. Therefore, automatically coding the medical record with standardized technologies is highly desired. In this traditional community generated systems, user generated system, which they employed has many limitations which are as follows.

1. *Trouble in replying* : Not all questions are answered by the doctors and the reply is based on the doctors availability. There is no instant response making patients to wait for experts acknowledgement for a long time.
2. *Lack of compatibility* : Some questions answered may contain medical terminologies and cannot be used by patients directly and there is no compatibility between them.
3. *Payment methods*: Fewer online services would charge amount to perform live chat or telephony communication with doctors online.

III. PROPOSED SYSTEM

Auto-generated system is a recent approach developed to overcome these problems. Chatbots is a computer program which conducts a conversation via auditory or textual methods. It interacts with human users. Automatic content generation systems are expected to be more reliable and compatible due to the presence of local medical coding and global learning approach. This system bridges the gap between health seekers and healthcare providers. Further this system

provides instant and relevant replies to the queries. At earliest stage this automatic mapping medical record system uses rule

based and machine learning approaches to code medical terminologies with standardized terminologies. Also this previous work used only the simple medical dictionary to code the medical records, which may bring inappropriate terminologies.

Our analysis led to propose a scheme that used to prune the irrelevant data of specific dataset, collect external authenticated vocabularies and map it to the medical concepts using two components namely, local mining and global learning.

The proposed system consists of:

  1.Input gathering and data pre-processing

  2. Medical terminology detection

  3. Mapping relevant document

  4. Generating answers and solutions.

INPUT GATHERING AND DATA PRE-PROCESSING

FIG 1: Use case Diagram - Input gathering and Data preprocessing

Description

In this module, the user enter query that concern their health. Then input query is then pre-processed by extracting noun phrase. This also eliminates the noisy and filler data. Noun phrase detection is done using parts of speech tagger. A fixed pattern of part of speech tags are searched and word/tag pairs matching the pattern are pulled out from the text as noun phrase.
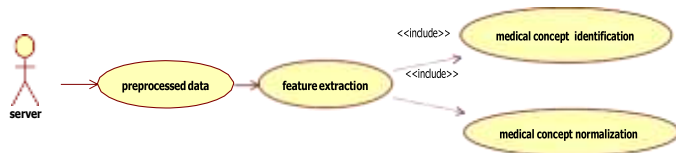
MEDICAL TERMINOLOGY DETECTION

FIG 2: Use case Diagram-Medical Terminology Detection

Description

This stage aims to differentiate the medical concept from other general noun phrases. The keywords are extracted by the frequency of occurrence using feature extraction. This module prunes the irrelevant data of specific data set. The medical

terms are then normalized in a standard form. Normalization is necessary because medical domain specific noun phrase cannot ensure that they are standardized terminologies.
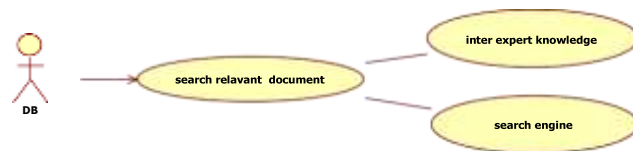
MAPPING RELEVANT DATA

FIG 3: Use case Diagram- Mapping relevant data

Description

The target of this section is to map the medical words with the record content present in the dataset using lexical similarity by graph-based global learning. Lexical similarity is a measure of degree to which the word sets of two given language are similar. Global learning works towards enhancing the local medical coding via collaboratively discovering missing key terminologies.
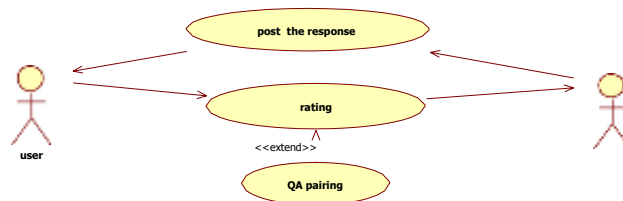
GENERATING SOLUTIO

FIG 4: Use case Diagram- Generating solution

Description

The relevant solutions will be posted as answers to the user query and the user is requested to rate for the response produced, based on which QA pairing would take place for later queries by different user.

IV.   ARCHITECTURE

Noise Elimination:

Initially, the user will login or the user is asked to register if he is a new user. Now the user is allowed to enter query regarding health issues.

The entered query is then preprocessed by extracting noun phrase. This is done by eliminating the noisy, filler and stemming words. Noun phrase extraction is done by using Stanford POS tagger. The sentence sequence must match the following pattern.

    (Adjective | Noun) *(Noun Preposition)? (Adjective | Noun)*Noun

(i.e.) the noun phrases should contain zero or more adjectives or nouns, followed by an optional group of a noun and a

preposition, followed again by zero or more adjectives or nouns, followed by a single noun.
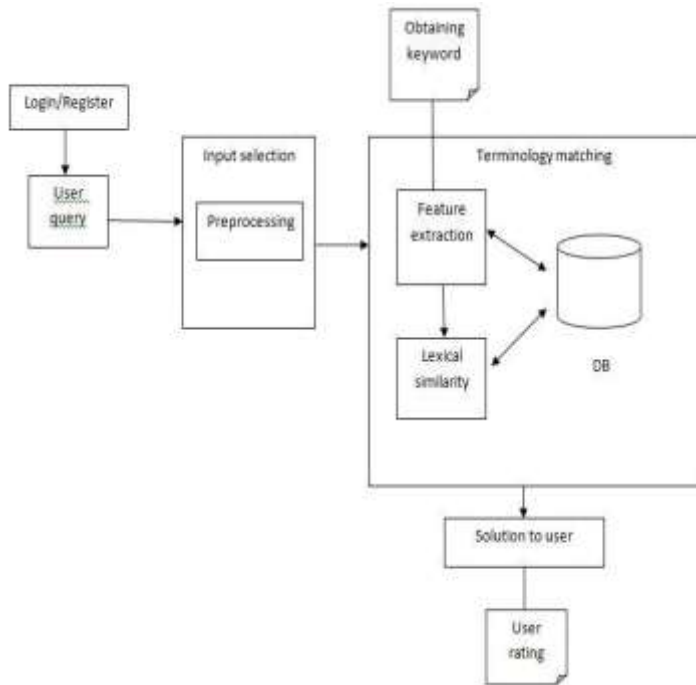

FIG 5: System Architecture

TERMINOLOGY MATCHING:

Feature Extraction:

The feature extraction contains two processes medical concept detection and medical concept normalization. The medical terms are detected from noun phrase by using medical concept detection. Although medical concepts are defined as medical domain- specific noun phrases, we cannot ensure that they are standardized terms. Hence we match the medical concepts with standardized terminologies in the database using medical concept normalization.

Lexical Similarities:

If all the keywords are mapped with the database, it falls under positive boundary. Then the appropriate solution is given to the user. If none of the keywords match the database then it is under negative boundary. In that case doctor's number will be provided.

Database:

Database contains the user details which are entered in the login credentials. When the user enters the query it is sent to the server for the operation, where the keywords extracted are matched as in the database. The results are immediately sent back to the user.

Evaluating the Solution:

Relevant keyword Selection process consists of extracting the medical terms from the user query and matching it with the dataset that could answer questions of the user easily. The solutions are extracted by applying a technique based on

comparing the medical keywords in the query and, those data sets that at least contain a keyword extracted and are evaluated by this answer extraction module that decides if it correctly answers the user question. If the extracted keyword doesn't present in our dataset, then the contact of specialist is provided. Now the user is asked to rate for the solution produced, based on which QA pairing would take place for later queries by different user.

V.   MODULES

SNAPSHOTS

1. Home page

2. Particulars page

VI.   CONCLUSION

In real world, people tend to seek knowledge from internet for every problem they face. This trend has increased not only for general topics but also their personal health concerns. In this paper we present a medical terminology assignment scheme to bridge the vocabulary gap between health seekers and health care knowledge. In existing system, a tri-stage framework is used to locally code each medical record. This local mining approach may suffer from information loss and low precision, which are caused by the absence of key medical concepts and presence of the irrelevant medical concepts. Global learning which provides a detailed description and outcomes the problem of local mining is used. It collaboratively learns and propagates terminologies among underlying connected medical records. It also enables the integration of heterogeneous information. Extensive evaluations on a real-

world data set demonstrate that our scheme is able to produce promising performance as compared to the prevailing coding methods. Even though these promising approaches exist, the system doesn't produce solutions in a well-sorganized manner. This motivates us to provide relevant solutions to the user by checking the boundary condition. Relevant keyword Selection process consists of extracting the medical terms from the user query and matching it with the dataset that could answer questions of the user easily. The solutions are extracted by applying a technique based on comparing the medical keywords in the query and, those data sets that at least contain a keyword extracted and are evaluated by the next answer extraction module that decides if it correctly answers the user question. If the extracted keyword doesn't present in our dataset, then the contact of specialist is provided. More importantly the whole process of our approach holds potential to handle large-scale data.

VII. FUTURE ENHANCEMENT

In the proposed system, we focus on giving appropriate answer for the health seekers if the user question contains medical terms which match our database. Else doctor's contact will be provided in case if none of the keywords match with our database. In the future, we will investigate how to flexibly organize the unstructured medical content into user needs-aware ontology by leveraging the recommended medical terminologies. Since many solutions are resulted in our system, we will enhance by asking the user to rate the solution and based on this we will do QA pairing. We can also incorporate the spell check of medical words and also try to improve the accuracy, speed receiving of answers.

VIII. REFERENCES

[1] L. Yves A., S. Lyudmila, and F. Carol, "Automating ICD-9-cm encoding using medical language processing: A feasibility study," in Proc. AMIA Annu. Symp., 2000, p. 1072.

[2] Z. Fu, G. Lu, K. M. Ting, and D. Zhang, "A survey of audio-based music classification and annotation," IEEE Trans. Multimedia, vol. 13, no. 2, pp. 303–319, Apr. 2011.

[3] L. Nie, Y.-L. Zhao, X. Wang, J. Shen, and T.-S. Chua, "Learning to recommend descriptive tags for questions in social forums," ACM Trans. Inf. Syst., vol. 32, no. 1, p. 5, 2014.

[4] M.-Y. Kim and R. Goebel, "Detection and normalization of medical terms using domain-specific term frequency and adaptive ranking," in Proc. IEEE Int. Conf. Inf. Technol. Appl. Biomed., 2010, pp. 1–5.

[5] L. Nie, M. Akbari, T. Li, and T.-S. Chua, "A joint local-global approach for medical terminology assignment," in Proc. Int. ACM SIGIR Conf., 2014.

[6] 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.

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

[8] 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.

[9] G. Zuccon, B. Koopman, A. Nguyen, D. Vickers, and L. Butt, "Exploiting medical hierarchies for concept-based information retrieval," in Proc. Australasian Document Comput. Symp., 2012, pp. 111–114.

[10] Y. Chen, Z. Chenqing, and K.-Y. Su, "A joint model to identify and align bilingual named entities," Comput. Linguistics, vol. 39, no. 2, pp. 229–266, 2013.

[11] P. Bansal, S. Bertels, T. Ewart, P. MacConnachie, and O. James, "Bridging the research–practice gap," Acad. Manag. Perspectives, vol. 26, pp. 73–91, 2012.

[12] N. Chu, Y. Choi, J. Wei, and A. Cheok, "Games bridging cultural communications," in Proc. IEEE Global Conf. Consumer Electron., 2012, pp. 329–332.

[13] M.-Y. Kim and R. Goebel, "Detection and normalization of medical terms using domain-specific term frequency and adaptive ranking," in Proc. IEEE Int. Conf. Inf. Technol. Appl. Biomed., 2010, pp. 1–5.