IMPLEMENTATION OF MEDICAL KNOWLEDGE EXTRACTION FROM TRUTH DISCOVERY AND SUPERVISED METHODS

S. Neelakandan[1] R.Annamalai[2] S.Velmurugan[3]

[1&2]Assistant Professor, Department of IT, Jeppiaar Institute of Technology.

[3]Assistant Professor, Department of CSE, VelTechMultitech Dr.Rangarajan Dr.Sakunthala Engineering college.

Abstract

The medical question and answer website plays a vital role in today's fast world because everything is becoming digital .When the world is stepping forward to become digital ,then storage of data becomes a complex task .If a Question and Answer website is considered for extracting medical data ,it is observed that each and every data that is generated becomes valuable. Thus, storing all these data by filtering out the unwanted data is a challenge .This is not possible when gathered information from the patients in a detailed manner and then separating it into useful information .When proposed a system to avoid the time taken ,cost of extraction of the useful information from unwanted sentences .Moreover ,our main aim is to reduce the volume of the data that is stored in the back end for faster retrieval of data. Here proposing an additional feature of leaving a message in chat to the doctors and whenever the registered users login to the account they can see details about any camp .This project has another advantage to give the patients some first aid related videos according to their wish .The implement this project in the name of, Easy Med which can automatically give you suggestions.

Index Terms – Knowledge Extraction, Supervised Methods, Trustworthiness

1.INTRODUCTION

As the world where we live runs very fast, all the traditional activities are in a revolutionary track. Besides following the traditional method of approaching a doctor personally and asking suggestions, online websites makes the work easy. A study says that there are hundreds of heath problem related queries that are searched every day in Search Engine. Globally the online health care websites are considered as a million-dollar industry. These websites have a million of registered users and doctors.

Compared to traditional one to one service given by doctors ,the online medical health websites has a very good future .For example *WebMD* alone receives about thousands of health related questions daily. There is no doubt that every data that is generated by using this website is a quality data [2]-[5],but how to minimize the volume of the data that is generated during this extraction is a complex task .If able to reduce the amount the data that is stored and also get some valuable information then this can be given to construction of a robot doctor who can suggest remedies for the questions that is raised by the patients.

One of the important challenge taken an online health care website for extracting medical field is to reduce the volume of the data stored in the database .This is because if a medical website is taken there will be many questions symptoms that is mentioned in the same question related to a particular disease .If a patient asks this question more elaborately in the website ,then the volume of data that is stored increases .The need to address this challenge by reducing the amount of data is stored in the database is raised. To reduce the volume of data, the need is to first differentiate between relevant and correct information. This is because the question that is asked by the patients may be correct in their point of view whereas in the database there will be some other data, but when seen to the question, it will be correct but not relevant.

To address this challenge, one possible solution is to give a template for asking the questions in the way what way the project wants. Instead of making the patients to type a large question, this project gives a template to make the patients easy to give answers. If this is carried over, information's are easily given to the patients and the time taken to retrieve answers from the database becomes a very easy task. The key

component of this medical data extraction system is to provide remedies for the questions without any supervision The information given should be very clear and confident as this is medical data where the use of statistical data for giving answers or suggestions .Some existing truth discovery models [6]-[11] assume that the answers from the expert are reliable and those give the answers for the certain questions should also be reliable .Next challenge is to reduce the amount of data stored and the transformation of unstructured data into structured data .For this transformation of unstructured data into structured data a method called *entity based representation [1]* is used.

Although entity based representation transforms the unstructured data into structured data, the amount of data that is stored in the database does not decrease. A given a solution to this challenge by giving a template for the patients which asks minimum questions to the patients and give the suggestions. By doing this reduction of amount of data that is stored in the backend. Another additional feature that is given to the registered users is to leave a message in chat. The doctors who are involved in giving this service have a separate login where when they login they can be seeing the message that is left for them. This will work only for simple diseases like fever, cold or some non-serious diseases. But if the person has a serious disease then first aid is needed at that time. For resolving this kind of analogies, there will be some health-related videos that are mentioned in the website. These videos are straightaway given by expert doctors so that the trustworthiness of each data is assured.

Our contributions towards this project are:

1.Reducing the volume of the data that is stored in the database by reducing the amount of words that is typed by the patient by giving template common to those who register in this application.

2.Creating a chat box for the patients to leave the questions and prescribing the patients about the doctors and the field of expertisation.

2. OVERVIEW OF THE SYSTEM

The objective of system is to reduce the amount of the data that is given by the patients. Because if the patient gives his question in a paragraph then each word must be checked in the medical dictionary and the perfect match must be found and then give the suggestion. If the patient is given a template where he needs only to enter his disease, it becomes easy to extract the details. The propose EasyMed system that can jointly perform the extraction of medicines and chat with the doctors in personal so that the patient can get a clear view about their disease. If it is a critical disease, then the videos that is given in the website may help them. Firstly, the patient gives their main disease as their question topic and this system searches for the prescriptions per this topic. Based on this question topic, the patient may also get some videos that is spoken by expert doctor that gives them a trustworthy first aid methods. This also give a place for the users to know about any free checkups or any camp when they login. The patient gives the question topic based on which the extraction of details takes place. After a while, considering their wish to get update or not and the question topic on which their search is done some free health checkups and donation camp is given.

3.METHODOLOGY:

In the proposed system, overcoming the problem of noisy data and handing the big challenge in big data, is storage by getting the exact information from the patients itself.

PROBLEM FORMULATION AND NOTATIONS:

Question topic: For each question given by the patient, assume that its related to the topic such as Pulmonology. This says the exact problem of the patients in a single word say Pulmonology. Directly keep such question information for the system.

Symptoms: Usually when a patient explain his problem he/she gives only the symptoms for his problem, but in our system when he/she gives the question topic, a few symptoms comes in the dropdown.

Admin: He is a person who looks care of the supervised database and includes the prescription for each problem given by the patients. Admin people also keeps the information of the Specialist doctor, Doctors contact number and address. He also gives information about the camps conducted in various places to the registered patients.

Doctor: a doctor is a person who answers questions on the medical Question and Answer websites. On the website from which the data is gathered, the “doctors” are real doctors, though it may not be this case for other websites.

Answer: an answer is the prescription which is given from supervised database. It may be multiple answers provided by different doctors for the same queries, and these answers may be noisy and unreliable in other websites. The prescription given from the database is already checked and the data is provided by the expert doctors in field.

4. CHALLENGES AND SOLUTIONS:

Although in existing system, Medical Knowledge Extraction is formulated, in our existing system we overcome some unique challenges of Big Data

Noisy Input:

The first challenge faced is how to clear the noisy input. In the existing system, an entity based representing method of text is used. Extract a set of entities from each question q ∈ Q to represent the original question , where an entity can be a symptom, disease, drug, etc. Correspondingly, the entities are looking for from the answer are disease, drug, drug side-effect and other problems. The available medical entity dictionary for entity extraction. If a word from a queries text exists in the dictionary, then put that word into the entity set for this type of question. As the age of patient is important for diagnosis, so it is mandatory to include the age information in the entity set. To overcome this system, directly get the problem of the patients as question topic and give the prescription for the patients. As the question topic is a single word, problem of the noisy data is overcome and when each question topic is given, the symptoms of the problem comes in a dropdown.

Data in Voice Format:

The second challenge faced is getting the data in a voice format. In the existing system, the patient's problem in the voice format the patient himself or with someone's help gives the disease in his own voice. If this is the case, there might be some problem if the disease name has been understood wrongly. There may be any hindrance if the voice is not correctly heard because every data is valuable. For example, if the patient has cold and cough and he needs to get some advice from the doctor then his voice may not be clear at that instance. To address this problem, get the information by typing the disease name.

Trustworthiness:

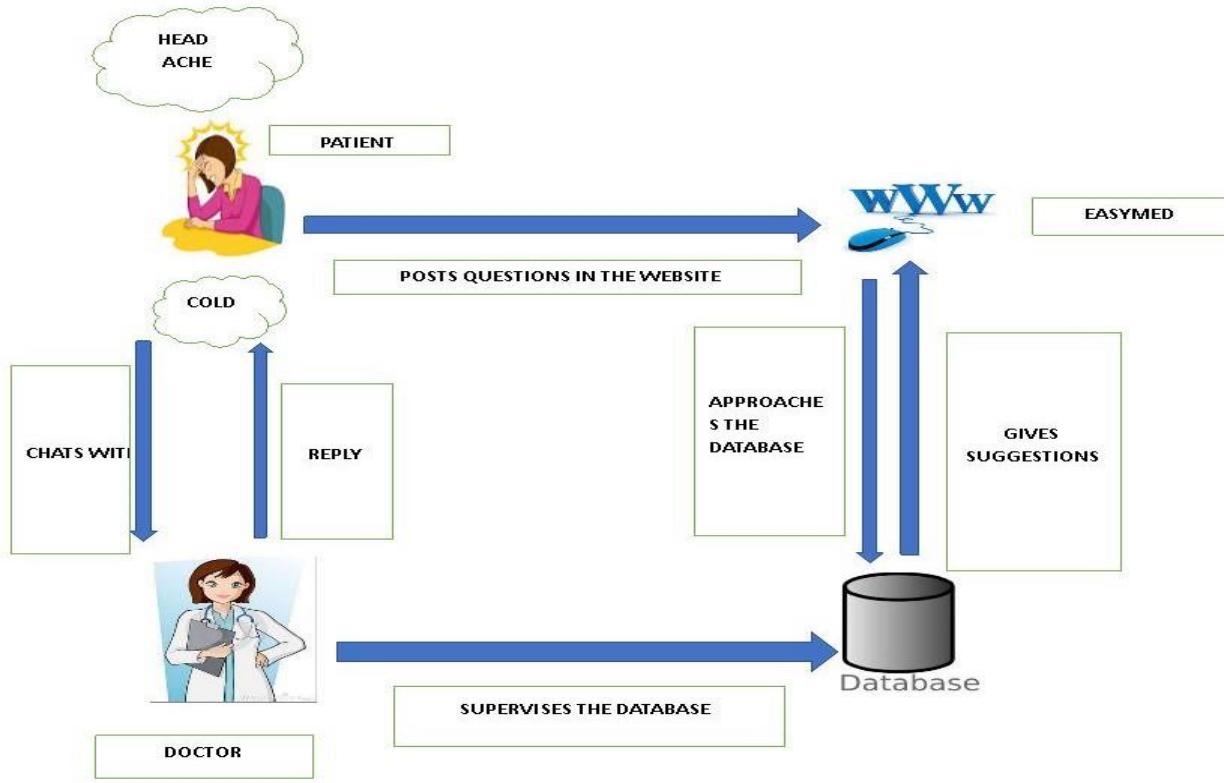
The data that is got from any website or database cannot be assured to be trustworthy. To assure this trustworthiness, give some videos that is delivered by the expert doctor. If the patients wished to get some related first aids, then the registered user use these videos by giving their correct disease as question topic**.**

Chatting Facility:

If the patient needs to contact the doctor in personal, two options are given. The first option is to chat with the doctor. The doctor after seeing the message that is posted, he/she gives his suggestion. The second option is to give their contact details and some more information about the doctor that includes his/her area of expert and experience.

5.IMPLEMENTATION

This project is implemented in three modules (Patient, Admin, Doctor). The patient gives his personal details when he wants to use this service. The patient gives his name, email id and mobile number when he/she sings up for this website. After all the criteria is met, the registered user can login to the page and get the suggestions. The patient after logging into the website, he/she first specifies the question topic where the patient gives the main disease he suffers from. For example, it may be migraine. After specifying the question topic as the main disease, the symptoms that are related to that disease is given in the symptom drop down box. Also, the patient specifies the medicine that is allergetic to them for giving correct medicines. Then the patient gets the suggestions and the doctor details for further contact.



The admin maintains the database which contains the symptoms, prescriptions, doctor name, doctor contact details and related videos. When the patients give all the details like question topic and symptoms, the database that is already created gives the correct prescription. If the patient needs to contact the doctor in personal, then the doctor's role comes into play. The doctor can chat with the patient by giving the prescriptions. If the patient needs some more information about the doctor, they can get the contact details from the database. If the user needs any first aid for any disease, then they can use the videos that is given in the website. These videos contain the doctor suggestions and the doctor gives his opinion on the symptoms.

6.RESULTS

The following Figures shows the how the doctors and patients interacted through chat and video materials. It showing the time duration and how many number of chat done between doctors and patients. It has implemented using R-Programming tool used for statistical computing also mining the data's.

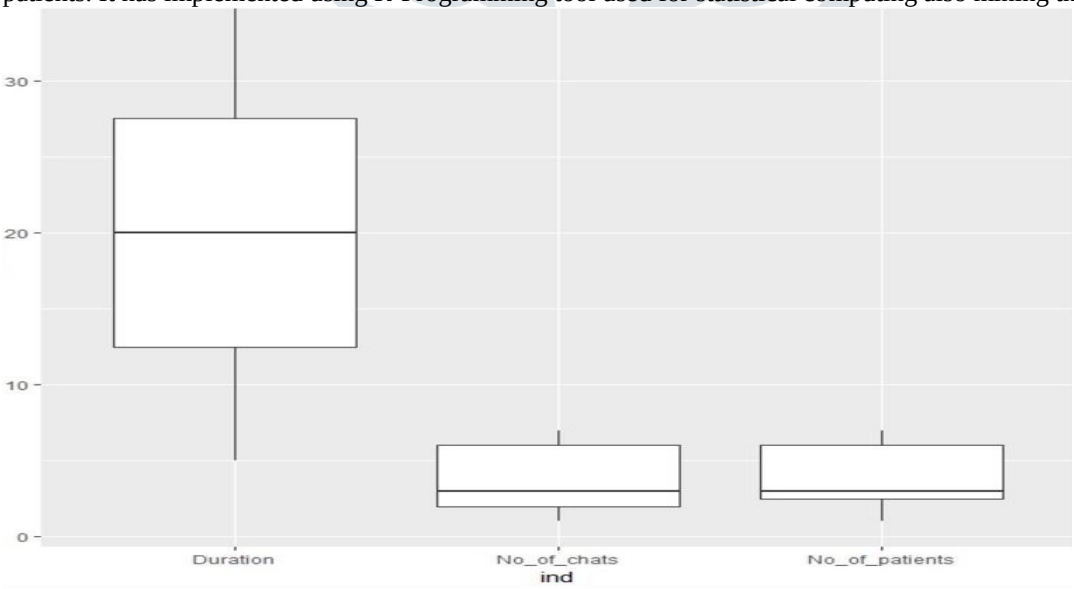


Figure 1 : Doctor Graph

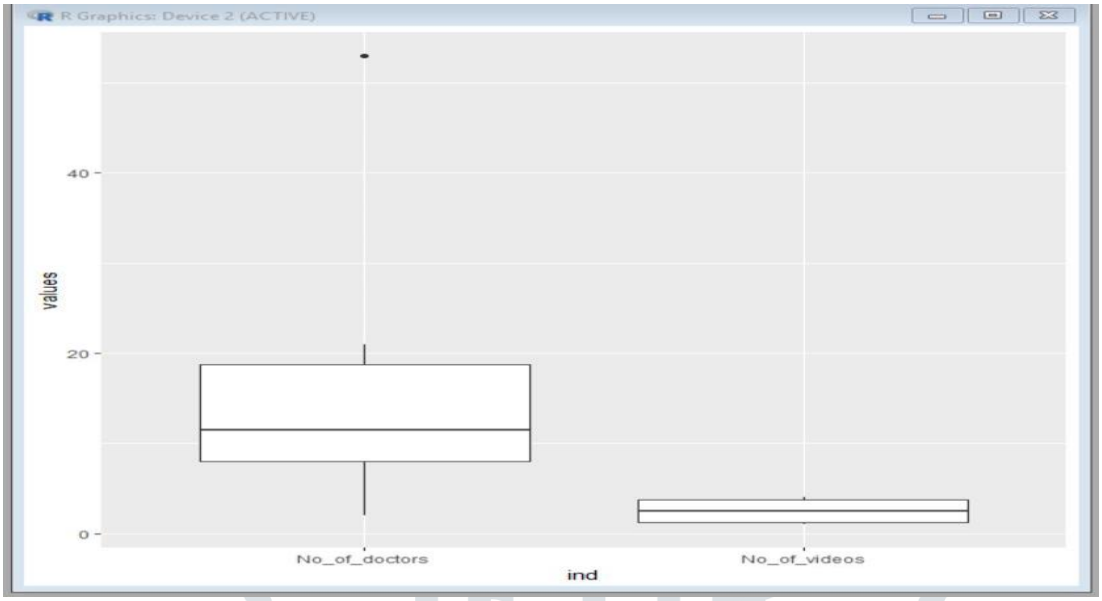


Figure 2 : Patient Graph

INPUT:
Question topic, Symptoms and the allergetic medicine
OUTPUT:
Prescriptions, contact details and chat facility.
STEPS:
1.Preprocessing of data and storing in the database.
2.Patients login.
3.Entering the question topic and the symptoms the patient experience.
4.Processing of the database.
5.Getting prescriptions and contact details

6.CONCLUSION

The data in the medical website is all valuable so we can't easily neglect those data. To get the correct data, we can directly ask the patients to enter the data which avoids more volume of data that is stored. This also reduces the time needed to extract the data. To ensure trustworthiness in the data, we give a chance to contact the doctors directly. For any first aid, related queries, we can get the suggested. We also show a real-world application, Easy Med, to demonstrate the impact of the MKE system. Beyond this website, the information's that are stored in the database can be given to the robots for future artificial intelligence. Based on these methods we have proved that the doctors can able to build a efficient communication with the patient also it will help to manage emerging situations. In future the communication process will be implemented in the major languages are used in the country however voice recognition process will be done to avoid many problems during input recognition.

REFERENCES

1. Yaliang Li,Chaochun Liu,Nan Du,Wei Fan,Qi Li,Jing Gao,Chenwei Zhang and Hao Wu,"Extracting Medical Knowledge from Crowdsoruced Question Answering Website,"IEEE Transactions on Big Data,vol 2,no. 2,Sep 2016.
2. L. Nie, Y.-L. Zhao, M. Akbari, J. Shen, and T.-S. Chua, "Bridging the vocabulary gap between health seekers and healthcare knowl-edge," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 2, pp. 396–409, 2015.
3. L. Nie, M. Wang, L. Zhang, S. Yan, B. Zhang, and T.-S. Chua, "Disease inference from health-related questions via sparse deep learning," IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 8, pp. 2107–2119, 2015.

4. L. Nie, T. Li, M. Akbari, J. Shen, and T.-S. Chua, "Wenzher: Comprehensive vertical search for healthcare domain," in Proc. of the ACM SIGIR International Conference on Research and Development in Information Retrieval (SIGIR'14), 2014, pp. 1245–1246.

5. L. Nie, M. Akbari, T. Li, and T.-S. Chua, "A joint local-global approach for medical terminology assignment," in SIGIR 2014 Workshop on Medical Information Retrieval, 2014, pp. 24–27.

6. X. Yin, J. Han, and P. S. Yu, "Truth discovery with multiple conflicting information providers on the web," in Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07), 2007, pp. 1048–1052.

7. X. L. Dong, L. Berti-Equille, and D. Srivastava, "Integrating con-flicting data: The role of source dependence," The Proceedings of the VLDB Endowment (PVLDB), vol. 2, no. 1, pp. 550–561, 2009.

8. J. Pasternack and D. Roth, "Knowing what to believe (when you already know something)," in Proc. of the International Conference on Computational Linguistics (COLING'10), 2010, pp. 877–885.

9. Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in Proc. of the ACM SIGMOD International Conference on Management of Data (SIGMOD'14), 2014, pp. 1187–1198.

10. Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han, "A survey on truth discovery," arrive preprint arXiv:1505.02463, 2015.

11. S. Harispe, A. Imoussaten, F. Trousset, and J. Montmain. On the consideration of a bring-to-mind model for computing the information content of concepts defined into ontologies. In Fuzzy Systems (FUZZ-IEEE), 2015 IEEE International Conference on, pages 1--8, Aug 2015.

12. S. Harispe, S. Ranwez, S. Janaqi, and J. Montmain. The semantic measures library and toolkit: fast computation of semantic similarity and relatedness using biomedical ontologies. Bioinformatics, 30(5):740--742, 2014.

13. Sebastien Harispe , Sylvie Ranwez , Stefan Janaqi , Jacky Montmain, Semantic Similarity from Natural Language and Ontology Analysis, Morgan & Claypool Publishers, 2015

14. Xin Luna Dong , Evgeniy Gabrilovich , Kevin Murphy , Van Dang , Wilko Horn , Camillo Lugaresi , Shaohua Sun , Wei Zhang, Knowledge-based trust: estimating the trustworthiness of web sources, Proceedings of the VLDB Endowment, v.8 n.9, p.938-949, May 2015 [doi>10.14778/2777598.2777603]

15. C. Miao, W. Jiang, L. Su, Y. Li, S. Guo, Z. Qin, H. Xiao, J. Gao, and K. Ren. Cloud-enabled privacy preserving truth discovery in crowd sensing systems. In Proc. of the ACM International Conference on Embedded Networked Sensor Systems (Sensys'15), 2015.

16. A. Marian and M. Wu. Corroborating information from web sources. IEEE Data Engineering Bulletin, 34(3):11–17, 2011.

17. G. Seni and J. F. Elder. Ensemble methods in data mining: improving accuracy through combining predictions. Synthesis Lectures on Data Mining and Knowledge Discovery, 2(1):1–126, 2010.

18. S. Wang, D. Wang, L. Su, L. Kaplan, and T. Abdelzaher. Towards cyber-physical systems in social spaces: The data reliability challenge. In Proc. of the IEEE Real-Time Systems Symposium (RTSS'14), pages 74–85, 2014.

19. M. Wu and A. Marian. A framework for corroborating answers from multiple web sources. Information Systems, 36(2):431–449, 2011.

20. Z. Zhao, J. Cheng, and W. Ng. Truth discovery in data streams: A single-pass probabilistic approach. In Proc. of the ACM Conference on Information and Knowledge Management (CIKM'14), pages 1589–1598, 2014.