# Recommendation of Doctors and Medicines Using Review Mining

**Problem Statement**

Health information needs are also changing the information seeking behavior and can be observed around the globe. Challenges faced by many people are looking online for health information regarding diseases, diagnoses and different treatments. If a recommendation system can be made for doctors and medicine while using review mining will save a lot of time. In this type of system, the user face problem in understanding the heterogeneous medical vocabulary as the users are laymen. User is confused because a large amount of medical information on different mediums are available.

In this research, a SVM (support vector machine), VSM (vector space model) based health recommendation system (HRS) is discussed for user that contain a web based assessment system using review mining. Expected outcome of this project is that a recommender system will be available that will suggest doctors and medicine to users. Scope of the project is that this recommendation system can be used for e-business systems like Amazon's suggest service for products. The idea behind recommender systems is to adapt to cope with the special requirements of the health domain related with users.

**Background**

Information retrieval(IR) is a special research field of computer science that emerged in the 1970's. It addresses the human need to automatically find or filter relevant text documents from a potentially huge collection of managed documents, for this the user has to write a query. This problem is also known as the classical information retrieval problem as it saves user time and work of manually searching the document collection. But problem with these type of IR systems is that they rely on word statistics regarding managed documents. Modern IR systems perform document ranking, which means that the quality of a match between a query Q and a managed document D is assessed via a similarity function. IR relies on term matching for document ranking which will increase the sc. Another approach is to do match between two terms from Q and D which is based on exact match which is too strict so to overcome this Stemming and lemmatizing, Query spell correction techniques are used. Recent trend in recommendation system are the health recommendation systems are used now[1] and also some mobile app based health recommendation systems are used now for recommendation of doctors and medicine.

**Methodology**

**Step 1: Data collection and dataset preparation**

This will involve collection of medical information artifacts from various sources like hospitals, discharge slips of patients and then preprocessing is applied on dataset which will remove all the unnecessary data.
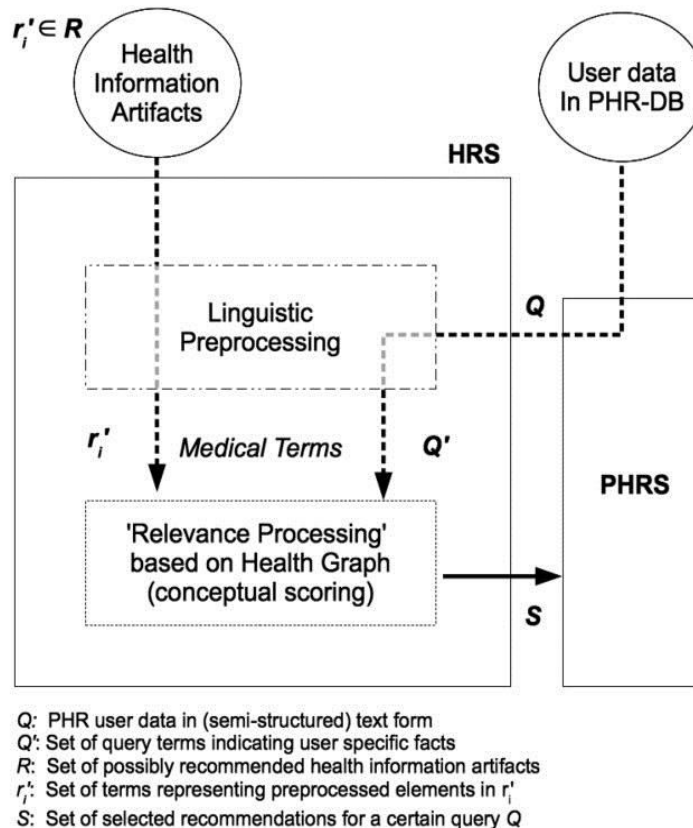
---

[1] http://www.fda.gov/Safety/MedWatch/

Q: PHR user data in (semi-structured) text form
Q': Set of query terms indicating user specific facts
R: Set of possibly recommended health information artifacts
$r_i'$: Set of terms representing preprocessed elements in $r_i'$
S: Set of selected recommendations for a certain query Q

Figure 1. Architecture of recommendation system for doctors and medicine[2].

## Step 2: Developing a SVM based model for recommendation of doctors and medicines using review mining.

In this step, a SVM (support vector machine) based advance health recommendation system (HRS) is developed as shown in Figure 1, also one VSD (vector space model) based naïve HRS is to be developed. A web based assessment system is to be made for recommendation of doctors and drugs where results can be checked at the time of deployment.

## Step 3: Training and experimentation on datasets

The recommendation model for doctors and medicine will be trained on both the dataset i.e., Heidelberg University Hospital that has 27,000 fully anonymized, real-world discharge letters dataset and 800 health information artifacts provided by the German Institute for Quality and Efficiency in Health Care.

## Step 4: Deployment and analysis on real life scenario

---

[2] Wiesner, Martin, and Daniel Pfeifer. "Health Recommender Systems: Concepts, Requirements, Technical Basics and Challenges." International Journal of Environmental Research and Public Health 11.3 (2014): 2580–2607. PMC. Web. 18 Apr. 2018.

The trained and tested recommendation model will be deployed in a real-life scenario by using a web based assessment system for further analysis where both naïve HRS, and advanced HRS will compete against each other in matching the recommendations made by the human experts & will be leveraged for further improvement in the methodology. It will also follow the proposed architecture shown in Figure 2 and its working is explained with the help of diagram.
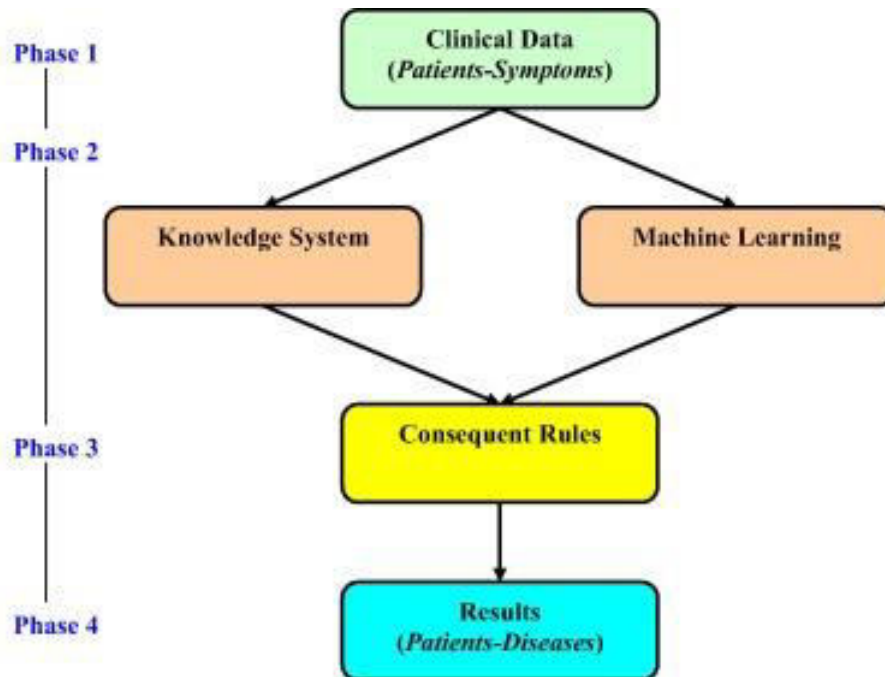
Figure 2. The recommendation system based on medical diagnosis process of the health care support system

**Experimental Design**

*Dataset*

Heidelberg University Hospital has 27,000 fully anonymized, real-world discharge letters dataset provided by them on request. Also MedlinePlus[3], Informed Health Online[4], WebMD[5] are available. Medical content i.e., 800 health information artifacts written especially for laymen is provided by the German Institute for Quality and Efficiency in Health Care IQWiG[6] which can be used for experimentation and evaluation.

---

[3] http://www.nlm.nih.gov/medlineplus
[4] http://www.informedhealthonline.org
[5] http://emedicine.medscape.com
[6] http://www.iqwig.de

*Evaluation Measures*

Measures such as accuracy, true-false ratio in terms of recommendations made after implementations of above system will compete against each other in matching the recommendations made by the human experts. Retrieval precision ($\rho$) of both systems will be measured and compared against each other & for the recommendation of doctors and medicines using review mining.

**Software and Hardware Requirements**

Python based Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs for training the SVM based naïve HRS, and advanced HRS.