

# Know Your Diseases!

Kashyap Bhansali  
Graduate Student  
Computer Science (MCS)  
Arizona State University  
Kashyap.Bhansali@asu.edu

Saloni Shah  
Graduate Student  
Computer Science (MCS)  
Arizona State University  
Saloni.Shah@asu.edu

Raj Dalvi  
Graduate Student  
Computer Science (MCS)  
Arizona State University  
rsdalvi@asu.edu

Shruti Mahajan  
Graduate Student  
Computer Science (MCS)  
Arizona State University  
Shruti.Mahajan@asu.edu

## ABSTRACT

Visual analytics is a kind of analytics in which the data is showcased visually displaying analytical results and insights extracted from large, complex dataset. Intelligent Visual Analytics is not only about showing statistics using charts but also involves text processing using text mining and Information Retrieval to represent important information visually. ‘Know your diseases!’ is one such kind of intelligent visual analytics. In this paper, we describe the analytical system in its entirety, ranging from the visualizations displayed and the principals involved behind its selection. The data from Q/A portal called WebMD is processed, cleaned and visualized to not only look aesthetically pleasing but also to convey the information visually. The goal of this system is to convey information to the fullest and consequently, encourage users to interact with the system to gain utmost knowledge about the diseases.

## Keywords

Health Data Visualization, Visual Analytics, Medical Analysis, WebMD, online discussion forums

## 1. INTRODUCTION

The essential essence of a visualization lies in effective narration of a tale. Vision is the most operative and powerful faculty to relay information from a large scale of data, which when represented by text or a data table, does convey a tale, but limits our perception of the information. It is important for intelligent visualizations to answer questions, support decisions and communicate information [1]. Today, one of the richest sources of information on public opinion are discussion forums found over the web. [2] Therefore, we have developed an Intelligent Visual Analytics for one such forum pertaining to health and nutrition, WebMD.

Over the years, an increase in health awareness has been realized and therefore there has been a significant growth in the number of users who have started using the rejoining and explanatory forums to gain more information related to health. [3] These forums facilitate easy and convenient solutions to audience’s query in a significantly less time. These systems allow people to interact with other users and share their knowledge related to health.

These forums also claim to solve problems of a share of population by answering just one individual’s question. Plethora of such web-forums, that provide remedies to health problems and particulars about diseases, symptoms and drugs, exists today, namely Patent Forum [4], eHealth Forum [5], EBIX A.D.A.M [6], etc.

WebMD is primarily an online publisher of news and information relating to human health and well-being. [7] It is widely used as a health forum inundated with a surfeit of symptoms, health concerns and expert advice. We’ve implemented an analytical system which covers a broad expanse of health topics, question-answers, WebMD community member details, and disease details. ‘Know your Diseases!’ lists the most prevalent topics and based on the topic selection, it displays an interactive visualization which shows the relationships between various diseases and symptoms, other related topics, the prevalence of disease over the past years, and reveals the affected region of human anatomy. Overall the idea is to allow the user to get a good overview of what is being discussed, how things are interleaved and also have the ability to get more details on demand from Q&A discussed on the WebMD forums. This analytical system aims at providing a complete analysis of a given disease topic discussed on the forum.

Section 2 describes the motivation behind this system. Section 3 discusses the visual design of the system followed by Section 4 discussing the implementation details. The paper is concluded by enlisting some unusual findings and discussing some future goals.

## 2. MOTIVATION

“Health is wealth” - There is nothing in our life that is more valuable than good health [8]. A healthy person is completely free from any illness or injury. Therefore, staying healthy and sound is very crucial.

The question and answer portals are modern form of an information exchange system. A study conducted by M. Dolores et al in [7] collects hundreds of definitional question from two health related websites WebMD and MedQA. The retrieved answers were tested for four evaluation measures like precision, mean reciprocal rank, total reciprocal rank and first hit success to check the quality of the answers given by users on this portal. It was seen that these systems are useful to retrieve basic healthcare

related information. Some of the answers from MedQA also give links to future research articles. This provides us a simple insight into how a question-answer forum can be a tool providing realm of information access.

Recently, the software firm Intel developed a way to provide personalized care [9] to individuals. The main idea behind this project was to stipulate unique health care information to a patient keeping in mind their situation and needs. They promise to help individuals to be proactive in deciding which patients are at high risk for readmission into hospital, admission to ICUs and similar other conditions. All of these are powered by their technology and health related data.

This served as a design inspiration for our analytical system ‘Know Your Diseases!’ which would assist in speculating health problems based on symptoms and would realize awareness of a disease in terms of its prevalence and related health issues. We planned to design and implement our Visual analytics system in a way to summarize the millions of Questions and Answers belonging to hundreds of topics discussed on WebMD forums.

We aim to provide answers to following research questions:

1. What kind of questions are asked on the forum? What are the most commonly discussed Topics & Diseases?
2. What are the Symptoms people with particular Disease have? What parts of the human body are affected by this Disease?
3. Prevalence of the Diseases in the population over the years?

Hence, a better health quotient was aimed at, while designing and implementing “Know Your Diseases!”

### 3. VISUALIZATION IMPLEMENTATION

‘Know Your Diseases!’ consists of interesting and informative visuals that helps us answer the research questions in an intuitive way.

It becomes important and challenging to use the right visuals and colors, because large amount of data is being processed and visualized.

We have selected our choice of visualizations based on the Visual Information seeking adage:

- Overview first
- Zoom & filter
- Details on demand

Following are the design principles that we follow:

- Correlation doesn’t imply causation
- Exaggerate or Simplify
- Generalizations improve clarity
- Emphasize important information or de-emphasize irrelevant details

It is important to have a way of revealing several levels of details, from a broad view to a fine structure. It is important in visual analytics to give the user a number of ideas in the shortest time with the least amount of ink in the smallest space [10] [11].

A right color choice improves the perception and makes understanding the data much more convenient. It is also necessary

to think how the whole interaction will simulate. In our implementation we have crosslinked the visualizations, i.e. we have coordinated these visualizations in a way where an interaction in one chart will drive the changes in all other charts to reflect that particular interaction.

This system analyses different groups of diseases into an interactive chart showing a hierarchy of health problems. Selecting one bubble subsequently alters the chord diagram to show related symptoms and health diagrams. This helps the user explore his related health problems which he may not be aware of and hence, helps him to analyse his situation better. To facilitate better understanding of diseases under consideration, our human physique highlights the human body area infected and the line chart explains how the disease has been widespread in the past, as per its discussion over the forum. This implementation provides a complete analysis of a disease from the WebMD dataset in its entirety, ranging from symptoms, related problems, and its prevalence in the world.

### 3.1 Data Collection and Processing

Initially, we started with the given dataset that included the details about the Topics, Questions, Answers, Participating Members and Related Topics. The basic idea to start with was to use text mining and information retrieval techniques to extract the important content.

First of all, it was important to identify the names of diseases and symptoms and drugs. Due to limited number of symptoms collected from the WebMD website we additionally extracted symptoms from another medical portal named Mayo Clinic. We used this scraped data to obtain relevant data from topics, question and answer data provided, to be used for our visualization system.

Text Mining or Information Retrieval involves many different techniques to extract important data from the dataset. For this system, we used the scraped data from WebMD and Mayo Clinic to draw out the relevant close matches from the questions and answers using Sequence Matching.

Sequence matching can be used to match sequences in strings which are hashable. This was preferred over TF-IDF because it reduced the occurrence of meaningless data and also took care of typos which could be there because of the complex spelling of diseases and symptoms. TF-IDF (term frequency-inverse document frequency) only looks for words in a corpus which it thinks are important but we are not just going for important words. All the data processing was don’t using multiple python libraries like Difflib, json, RAKE (Rapid Automatic Keyword Extraction), NLTK.

Other ways we tried extracting is using a combination of NLTK toolkit (a popular Natural Language processing toolkit for python) with RAKE (Rapid Automatic Keyword Extraction). We partially processed the data with these libraries and in the end settled with the functionality provided by the Difflib library to find the closest matches from the data.

With a bag-of-words approach, some clustering was performed using some manual inputs. A hierarchical structure, was then formed, which has been used to construct a zoomable bubble chart which is discussed in the next section.

All of the charts, are developed in D3.js, while we used most of the existing open source functionality, we modified the charts to link them together. We used the following technologies for

developing the visual analytics system: Python, JavaScript, HTML, CSS, SVG, JSON, JQuery.

### 3.2 Zoomable Bubble Chart

Zoomable bubble charts are basically a big blue bubble that has nested bubbles inside it. This sort of visualization is artistic and quite intuitive when we need to demonstrate a hierarchical structure in data. As the name suggests, this kind of chart will have multiple levels to zoom in. The nature of the bubble also provides one more dimension, that is, the radius of the bubble, which can be used to determine some quantity.

Using the same idea, we decided to use a zoomable bubble chart to represent a hierarchy of topics discussed on the WebMD forums.

We clustered a set of topics to build a hierarchy which would help the user to get a good overview and also get to the smallest detail if required. This chart is the driver for the rest of our visual analytics system.

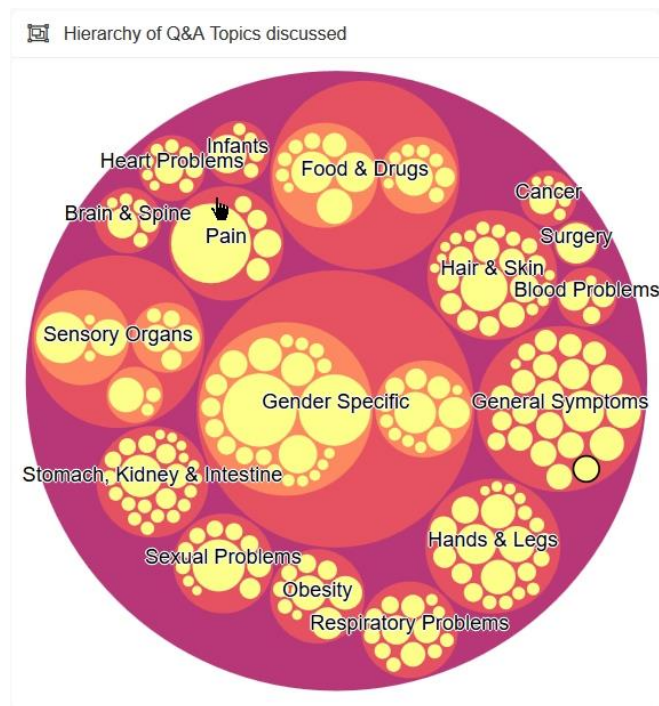


Figure 1. Zoomable Bubble Chart

The bubble chart implementation provides a wide taxonomy [25] of health problems grouped into following categories:

- Gender Specific.
- Infants
- Pain
- Hair and Skin
- Drugs
- Hands and legs
- Obesity
- Respiratory Problems
- Food and Drugs
- Sensory Organs
- Stomach, Kidney and Intestines
- Heart Problems

- Sexual Problems
- Surgery
- Blood Problems
- Brain and Spine

The groups have been intentionally named in layman terms so as to facilitate effortless selection. Each big bubble encompasses smaller bubbles, which zoom in on the selection of Parent bubble. Size of the bubble symbolizes the relative number of health complications it categorizes, or the number of.

The motivation of choosing a prescriptive bubble chart for taxonomy was to maintain uniformity in parent and child (representing all the diseases in form of bubbles one inside the other), and to provide easy navigation from one group of diseases to the other.

The radius of the bubble indicates which topics have the most number of discussions. This could help the user to get an overview as to what kind of topics are discussed and their occurrence level on the forum. It makes it easy to drill down to see comparison between individual topics belonging to the same cluster.

Color choices have been made taking into consideration the visualization principles of color and vision. A clear sequential scheme has been used to color the bubbles, cyan, blue and white (going down the category tree). [4] White color has been used to enhance visualization of the sub-most category in each parent category. Shades of blue have been used because blue symbolizes trust, peace and comfort.

For example, by only looking at the Zoomable bubble chart, we can see and tell the following things:

- there is a lot of discussion about Women health problems accompanied by Q&A's for General Symptoms like headache, nausea, burns, etc.
- Pain and infection are the most common complications associated with a gamut of diseases.
- Complicated issue like Cancer and Surgery have been less discussed and talked about on the forum, since they're critical enough to directly consult a doctor in-person.

### 3.3 Chord Diagram

Chord Diagram is a graphical method of displaying the relationships in data. The data is arranged radially on the periphery of a circle with the related data connected by means of arcs [12]. This kind of a chart was popularized in 2007 by a post on The New York Times that discussed "Close-Ups of the Genome, Species by Species by Species" [13].

An intelligent chord diagram demonstrates how symptoms and health problems can be linked to other health problems, i.e. the problems which can be simultaneously present or in other words, prevalence of which can make the present symptom into consideration, more probable. All the datasets, especially questions and answer sets have been analyzed to extract all the relations between a gamut of health issues. Figure below is a sample of chord diagram when smaller bubble muscle pain in parent bubble 'Pain' is selected.

Nature has created our body in a way that all the organs work together in a highly coupled fashion. A little malfunctioning in one, results in a complication in other system. For example, a small infection in esophagus, if left unattended, might result in

ulcers of stomach infection. Or in case one has a throat infection, then it may increase the probability of having ulcers. Therefore, for an analysis of a disease, it is very important to obtain related health problems, since it opens new avenues to understand the disease and to find ways to treat it, since treating one problem may help cure the other. Also, in the diagram, a certain chord connects all the problems which have occurred simultaneously together in some other instances. Different contrasting color options have been selected to display related yet different problems [10].

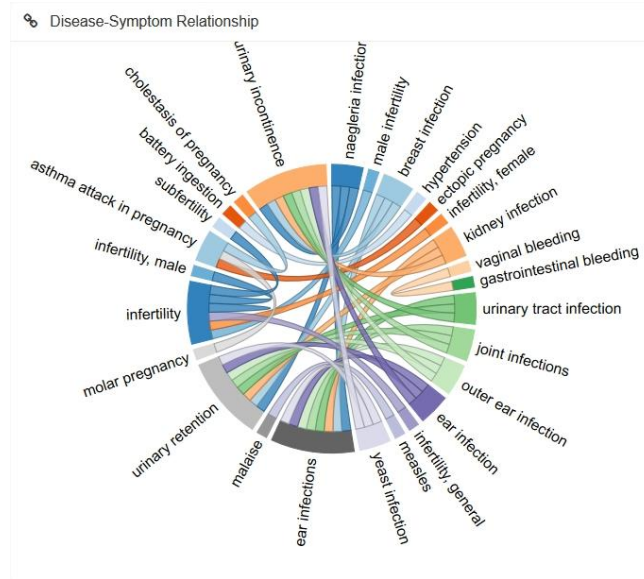


Figure 2. Chord Diagram

### 3.4 Line Chart

A line chart is a chart that plots a series of data points, connected by straight line segments [14]. A line chart is often used to visualize a trend in data over an interval of time [15]. Specially in case of the time series data. Thus the line is often drawn in a chronologically order.

In our project we make use of a simple line chart to demonstrate a trend in the prevalence of disease over the years or months. When a topic is selected, this line chart is updated based on the number of questions asked on the forum related to that particular topic over months or years.

A user using our intelligent analytics system could quickly answer if the questions related to certain topics are asked more or less and at around what time of the year. This could help in identifying the trend in symptoms that occur due to change in weather and seasons [26].

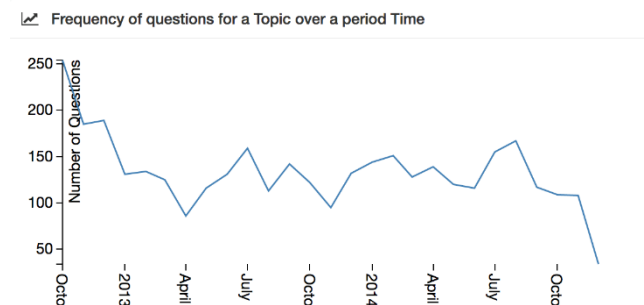


Figure 3. Line Chart

### 3.5 Human Physique

Knowing symptoms, related health problems or broad classification of your problem is not enough if you want to treat your health complication. The most important thing lies in the awareness of the anatomy part where the root of problem resides. For example, patches or rashes in skin, can be caused by the disease dengue, which is a viral attack, most importantly not on the outer visible skin area.

It would be a little unfair to assume that the user of this system would know about all the diseases and symptoms and what areas of the body internally or externally it affects. This encouraged us to add Human Physique diagram to the analytics system which gives a general overview of what areas might be affected with a particular disease.

This particular visualization displays the region in body where the selected bubble effects and realizes pain.

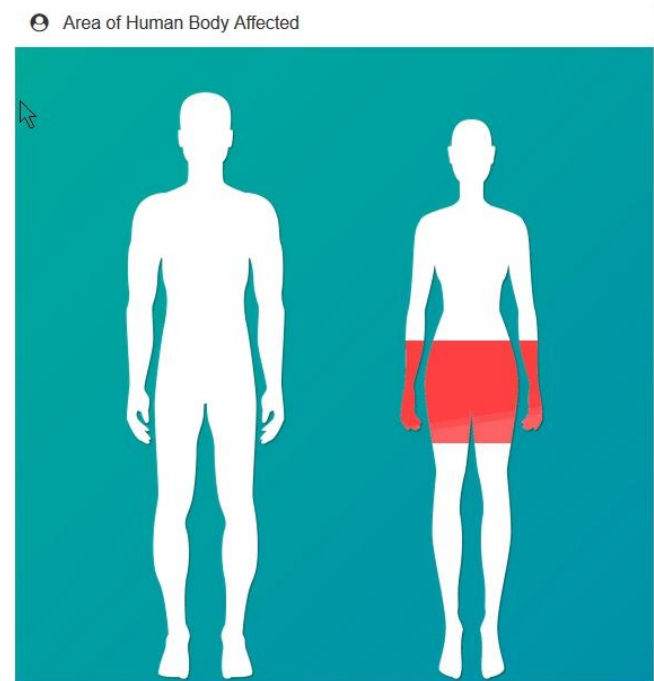


Figure 4. Human Physique

### 3.6 Relevant Questions and Medical Tests

Should knowing about a disease with respect to its symptoms, related health problems and prevalence be enough? Yes! But what can be better, than also displaying relevant top questions asked! This section of the analytics dashboard does the same. It displays the most recent questions asked related to a particular topic. The answers to these questions can be read by clicking on it. The question links directly to the WebMD website page where this question was asked, along with extra details like Name of the member posting this question, User ID, question post-date, etc. If this member has updated his contact info, he may be contacted by the info-seeking individual who would want to talk to someone who would already have gone through this particular health problem.

Therefore, this section sums up entire analysis of a disease by furnishing a direct instance of a disease into consideration. This section thus sums up all the analysis into a link which directs to an

occurrence of the disease. It provides a deeper understanding of the same.

Along with symptoms and drugs, we see that relevant medical tests also play an important role in providing information to users. Thus, in our visualization we mention the most relevant medical tests that are discussed amongst the people in question and answers.

#### Most Famous Questions

|  |
|--|
| How does alcohol affect pregnancy?             |
| How can lupus affect pregnancy or the newborn? |
| Is it safe to drink during pregnancy?          |

**Figure 5. Relevant Questions**

#### Medical Tests discussed

|   |
|---|
| ultrasound, pregnancy                               |
| tooth extraction for gum disease                    |
| scan, cardiac blood pool                            |
| amniocentesis for rh sensitization during pregnancy |
| oral glucose tolerance test                         |
| pregnancy test                                      |

**Figure 6. Medical Tests**

## 4. METHODOLOGY

The aim of our Intelligent Visual Analytics System is to help the user of the system to understand the data and answer as many questions as possible.

Our system was developed with the intension to answer the following research questions that have been mentioned earlier. They are:

1. What kind of questions are asked on the forum? What are the most commonly discussed Topics & Diseases?
2. What are the Symptoms people with particular Disease have? What parts of the human body are affected by this Disease?
3. Prevalence of the Diseases in the population over the years?

We discuss how the above questions are answered in the order they can be seen on our application.

To answer what kind of questions or topics are discussed on the forum, we thought of Clustering the topics based on similarity.

The bubble chart has various topics and on clicking any of the child bubbles, we see the diseases and connected symptoms in a chord diagram. The bubbles are generalized into topics to improve overall clarity [16]. The size of bubble indicates how commonly or rarely certain topics are discussed. For instance, there are maximum number of questions and answers related to Women Health problems followed by Sensory Organs in human body and general symptoms. This helps the user of the system to get a detailed overview of most commonly discussed topics and diseases. The zoom and filter capability help the user to know about the disease and symptoms in the linked chord diagram. More details are provided as the user zooms in to the innermost level of the bubble chart. Thus, the visualization provides complete justice to Visual Information Seeking Mantra taught in class [17].

Next we answer what sort of symptoms are related to a particular disease with the help of the Chord Diagram. Chord diagram as discussed earlier are the best for demonstrating relationships. A particularly complex dataset like this can be represented very well using a chord diagram. We map the relationships between the symptoms and diseases by identifying their occurrence in each question and answer of a particular topic. The aim in designing the chord diagram is to make large data sets coherent. The chord has multiple colors to show different categories of diseases and a disease and its symptoms are linked together with the same color. Ultimately, as discussed in class, the chord diagram conveys multiple complex ideas with clarity, precision and efficiency. In some cases, the data is not large enough to show a chord diagram. An instance where this helps is, when the topic pregnancy is selected, we can hover over “infertility” to see how it is linked to other entities in the chord diagram. This basically, helps the user to understand what sort of things were discussed in conjunction with each other.

To make a user feel more informed about certain topics and diseases which they might not know of, we use the Human Physique to demonstrate what area of the body in a male and female is affected. Majority of the body parts were extracted from the dataset but in some cases, manual inputs were needed to improve the accuracy. Another challenge in this human body usage was to highlight diseases such as diabetes which are not exactly centered on a particular organ of the human body. In that case, the whole body is highlighted or the closest possible body part is highlighted. When a bubble in the chart is clicked on, the chord diagram displays the diseases and symptoms. Apart from that, the affected body part is highlighted in the image of a human body.



**Figure 7. Overview of our System**



To answer the next question, where we give the analysis about the prevalence of certain disease topics we used a simple line chart to demonstrate a trend. This helps the user to understand a trend in the discussion of certain topics over a period of time.

Although our system was developed so as to convey the information discussed so far, it could be overwhelming for a new user. To support the information that could be obtained from the charts discussed till now, we also demonstrate a set of most relevant questions belonging to a topic for the user to understand it better. These questions are hyperlinks so that the user can navigate to the question thread on WebMD.

To further make use of the intelligent way to extract the data, we also give some additional information about the medical tests that have been discussed with each topic belonging to certain diseases.

Thus, in a way we try to answer all the questions that we thought of answering before and while developing this intelligent visual analytics system.

## 5. EVALUATION PLAN

Visual analytics and data visualization are designed for clarity in perception, therefore, we focus on visual interfaces and presentations, which are well understood, based on human biases and on how users interact with computers [18]. Our results are completely based on the questions and answers posted by humans on this forum. These humans are variety of groups ranging from patients, physicians to medical experts. Our results not only will help common people to gain basic knowledge but also health care professionals as they use symptoms and signs as clues that can help determine the most likely diagnosis for a disease. However, proper testing performed by a doctor is recommended to render a diagnosis because many patients experience some symptoms and not others [19].

We apply bottom-up method of processing as discussed in [20]. Firstly, a user extracts information just by having a glance at the analytics dashboard and knowing it showcases medical information without any other thought. The efficiency of this process is related to the large amount of neurons, up to five billion, which are simultaneously processing different features

## 6. DISCUSSION & FUTURE WORK

After successful implementation of the visual analytics, we see that some critical disease like Cancer, Blood problem and the likes have been discussed in less amount as compared to common and minor diseases. This indicates that users do not prefer to discuss critical topics on a forum and would rather choose seeing a doctor in person. We also observed that there is a lot of discussion about Women Health problems followed by Sensory Organs of human body. Pain and infections are the most common complication associated with a gamut of diseases.

After the implementation of showing how diseases and symptoms were related and their behavior over time, we also ought to display the areas in the world affected with these diseases majorly. For this, we dug into the dataset to find data related to geographical location of a user. In the near future, we would add a map visualization showing the occurrence of disease in various parts of the world. We would use choropleth maps showing the country highly affected by the disease in darkest shade and lightening its shade with countries affected relatively less.

This project can be extended into a visual recommender system for health related dataset. After the user enters a symptom, the system would recommend the diseases which commonly see that symptom and vice versa as well. Also, the commonly occurring diseases based on the users' geographical location would be displayed.

A series of dataset regarding medicines can be collected, hence visualized and consequently recommended by the system. There are so many pharmaceutical companies that manufacture a same kind of drug. Some medicine produces are instant relief providers and some provide treatment with time. Therefore, an expert/physician has vast amount of choices of medicines to recommend and prescribe. The proposed recommender system can recommend medicines that a doctor can remotely prescribe to a patient, based on the availability of medicine at patient's location and on how effective that certain drug is, at level of infection or pain.

While discovering about health concerns and related information, we came across Traditional Chinese Medicine (TCM). [21] Our research then made this interesting revelation that TCM has a very effective cure for most of the diseases and health problems. Their traditional and cultural medicines are very relaxing and relief providing. It can be really beneficial for ailing patients to consume this treatment. We aim at popularizing this idea in our recommender system by doing a little more research over this area and suggesting patients to use this medication if present. This might prove very helpful to the ailing patient.

Another idea we would to be emphasized is accessing the patient's location and recommending nearby doctors and physician. There are web services [22] available to do the same, but in limited areas. This idea could be embedded into our 'Know your disease' and hence a complete essence of a recommender system can be imbibed. What can be better than a system which presents a complete analysis of your system, recommend home remedies, give the user an example of a patient suffered/suffering from the same health problem by means of showing top discussion post, and also recommending to the patient, nearby doctors and physicians.

Why make a system only for the diseased? A normal healthy person should also ensure and maintain his health. A recommending and analytic system should also accompany articles related to health care and healthy lifestyle. An entire section suggesting health and wellness tips and articles can be incorporated into the system. Websites such as Health [23], MedicineNet.com, etc. can be embedded. There are so many minor problems a healthy person overlooks. This ignorance may someday cause a big health concern. Therefore, maintaining health is very important. 'Prevention is better than cure' is what we believe in. An occurrence of chronic fever or a lower abdominal may be a sign of UTI [24]. Initially trivial, this may lead a way for kidney stones or kidney infections which is lethal. Therefore, utmost care is essential.

We might include articles and tips suggesting or recommending exercises to maintain body fitness as well. Anyone may stumble upon some idea that maybe lead to his better health and hence be beneficial.

This is how a simple analytical system 'Know your diseases' can be extended in so many multifaceted ways and aid humanity towards better health and fitness regime.

After being acquainted with the basic principles of data visualization and analysis, we would be able to do complete justice to a recommender system for health concerns. Afterall, health is a very critical topic and should be carefully examined and treated.

## 7. ACKNOWLEDGMENTS

We thank our professor Dr. Sharon Hsiao for her reviews about our ideas and guidance in this project. Her presentation slides and weekly reading of research papers were very helpful in forming a basis for this project. We would also like to thank Arizona State University's School of Computing, Informatics and Decision Systems Engineering for providing us with an opportunity to learn this course of Data Visualization and broaden our horizons and ultimately, improve our skill set.

## 8. REFERENCES

- [1] White Paper: Principles of Data Visualization – What We See in a Visual.  
<http://www.fusioncharts.com/whitepapers/downloads/Principles-of-Data-Visualization.pdf>
- [2] Mitja T. et Al. Visualization on Online Discussion Forums. Technical Report. University of Maryland at College Park In JMLR; Workshop and Conference Proceedings, 2010.  
<http://jmlr.org/proceedings/papers/v11/trampus10a/trampus10a.pdf>
- [3] Maria-Dolares et Al. Question-answering systems as efficient sources of terminological information: an evaluation in Health Information & Libraries Journal, December 2010.  
<http://onlinelibrary.wiley.com/doi/10.1111/j.1471-1842.2010.00896.x/full>
- [4] <http://www.thepatientforum.com/>
- [5] [www.ehealthforum.com](http://www.ehealthforum.com).
- [6] <http://www.healthcaremagic.com/>
- [7] <https://en.wikipedia.org/wiki/WebMD>
- [8] Ankita Mitra. An Article: Short paragraph on Health is Wealth in Important India.  
<http://www.importantindia.com/8847/short-paragraph-health-wealth/>
- [9] An article: Intel Inside. More Personalized Care Outside.  
[http://www.intel.com/content/www/us/en/healthcare-it/personalized-care.html?cid=sem43700013298021834&intel\\_term=analytics+for+healthcare&gclid=CjwKEAiA1ITCBRDO-oLA-q\\_n8xYSJADjBQfGbJOcLgTZvW-myQ7jV4eL-z9osShcSHXssfRMSI9mDBoCbQzw\\_wcB&gclidsrc=aw.dsv](http://www.intel.com/content/www/us/en/healthcare-it/personalized-care.html?cid=sem43700013298021834&intel_term=analytics+for+healthcare&gclid=CjwKEAiA1ITCBRDO-oLA-q_n8xYSJADjBQfGbJOcLgTZvW-myQ7jV4eL-z9osShcSHXssfRMSI9mDBoCbQzw_wcB&gclidsrc=aw.dsv)
- [10] William C. et Al. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods in Journal of the American Statistical Association.  
<http://www.tandfonline.com/doi/abs/10.1080/01621459.1984.10478080>
- [11] Hadley Wickham. Graphical criticism: some historical notes in Journal of Computational and Graphical Statistics, 2012.  
<http://vita.had.co.nz/papers/stat-graph-hist.html>
- [12] [https://en.wikipedia.org/wiki/Chord\\_diagram](https://en.wikipedia.org/wiki/Chord_diagram)
- [13] An article: Close-Ups of the Genome, Species by Species by Species.  
[http://www.nytimes.com/imagepages/2007/01/22/science/20070123\\_SCI\\_ILLO.html](http://www.nytimes.com/imagepages/2007/01/22/science/20070123_SCI_ILLO.html)
- [14] Burton G. Andreas (1965). Experimental psychology. p.186.
- [15] Neil J. Salkind (2006). Statistics for People who (think They) Hate Statistics: The Excel Edition. page 106.
- [16] Tufte. Data-Ink and Graphical Redesign. In The Visual Display of Quantitative Information. 2010
- [17] Agarwal et Al. Design principles for visual communication. in ACM (2011), 54(4), 60-69. doi: 10.1145/1924421.1924439
- [18] An article: 7 Tips for Evaluating Data Visualization Software. <http://data-informed.com/7-tips-for-evaluating-data-visualization-software/>
- [19] An Article: Difference between Symptom and Sign.  
[http://www.medicinenet.com/symptoms\\_and\\_signs/article.htm](http://www.medicinenet.com/symptoms_and_signs/article.htm)
- [20] Jacqueline Strecker et Al. Data Visualization In Review.  
<http://idl-bnc.idrc.ca/dspace/bitstream/10625/49286/1/IDL-49286.pdf>
- [21] <https://nccih.nih.gov/health/whatiscam/chinesemed.htm>
- [22] <https://www.practo.com/>
- [23] <http://www.health.com/health/gallery/0,,20513624,00.html>
- [24] An Article: Urinary Tract Infection (UTI)  
<http://www.mayoclinic.org/diseases-conditions/urinary-tract-infection/basics/definition/con-20037892>
- [25] Shneiderman, B. The eyes have it: a task by data type taxonomy for information visualizations (1996)
- [26] McLachlan et al. LiveRAC: interactive visual exploration of system management time-series data in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2008, pp 1483- 1492