IMPACT OF SOCIAL NETWORKING SITES ON HEALTHCARE

BY:

SHIVANGI BOHRA (101513)

Under the supervision of:

DR. DIPANKAR SENGUPTA

May 2014

Submitted in partial fulfilment of the Degree of

BACHELOR OF TECHNOLOGY

In

BIOINFORMATICS

DEPARTMENT OF BIOTECHNOLOGY AND BIOINFORMATICS

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY,

WAKNAGHAT

Table of Contents

CERTIFICATE

This is to certify that the work entitled "**Impact of Social Networking Sites on Healthcare**" submitted by **Shivangi Bohra (101513)** to the Department of Biotechnology and Bioinformatics, Jaypee University of Information Technology, Waknaghat in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology in Bioinformatics** is a record of bona fide research work carried out by her under my supervision. This work has not been submitted partially or wholly to any other University or Institute for the award of this or any other degree or diploma.

Dr. Dipankar Sengupta
Assistant Professor (Grade-I)
Department of Biotechnology & Bioinformatics
Jaypee University of Information Technology
Waknaghat-173234
Solan, Himachal Pradesh

ACKNOWLEDGEMENT

I would like to express my deep sense of gratitude and appreciation to all the people who have contributed in this project. First of all I would like to thank my Project Guide **Dr. Dipankar Sengupta** for providing me with such a great opportunity of working on this project, having excellent future prospects.

I would also like to thank **Prof. R.S.Chauhan**, Head of Department, Department of Biotechnology and Bioinformatics, for providing me with the facilities to carry out this project. I would also like to thank **Foundation for Advancement of Education and Research** for selection of my project work under the **Motorola Scholar Program 2013-14**.

I am indebted to other faculty members, my brother and all those who provided their reviews and suggestions for improving this project.

Shivangi Bohra (101513)

List of Figures

List of Tables

Abstract of the Dissertation

Social media has moved beyond being a tool for individuals especially the young generation to share their private lives (pictures, messages) to fostering serious discussions on technology and business. Increasingly driven by regulatory pressures, the need to "*get it right the first time*" and minimize costs remain a concern in the healthcare industry. Customer feedback to improve business has thus become very important. User generated content in the form of peer reviews on service/products often paves the way for business to understand any unique requirements as well as the pain points of the existing services they provide. Thus, social media is making interactions between end users and service providers possible by providing relatively simple, easy to access and unbiased platforms for sharing feedback.

With the rapid growth of online social networking for health, health care systems are experiencing an inescapable increase in complexity. This is not necessarily a drawback; self-organising, adaptive networks could become central to future health care delivery. Patients are using social networking to access and contribute health information. Among those living with chronic illness and disability and engaging with social networks, there is considerable expertise in assessing, combining and exploiting information. Social networking is providing a new landscape for patients to assemble health information, relatively free from the constraints of traditional health care. However, health information from social networks currently complements traditional sources rather than substituting for them. Networking among health care provider organizations is enabling greater exploitation of health information for health care planning. The platforms of interaction are also changing. Patient-doctor encounters are now more permeable to influence from social networks and professional networks. Diffuse and temporary platforms of interaction enable discourse between patients and professionals, and include platforms controlled by patients. We argue that social networking has the potential to change patterns of health inequalities and access to health care, alter the stability of health care provision and lead to a reformulation of the role of health professionals. Further research is needed to understand how network structure combined with its dynamics will affect the flow of information and potentially the allocation of health care resources.

The power of social media networking is vast. Sharing thoughts, ideas, viewpoints, posting updates, collaborating with consumers and colleagues is immeasurable. Tapping into a community of users whose word-of-mouth influence in the social space is fierce, and it goes beyond the standard role of social media networking. Facebook, Twitter and Google plus are only a few of the social networking

platforms utilized, and the millions of individuals who use it have the capability to spread information like wildfire. They can reach and influence others in their social circles at lightning speed. Individuals have the capability to influence their friends about their favorite restaurant, movies, electronics and TV shows; but imagine the power that individuals have to influence their circle of friends, and their friends and so on and so on, about better health. Within the circles of social networking, trust and relationships are formed. Individuals can take an active role in promoting health and wellness. Social influence develops based on the trust within the circles of the social network. Family and friends can help inspire and motivate each other. They can also hold others within their circles accountable for their actions. People with influence and trust can help others achieve their health goals.

A major problem being faced by end-users while looking into the social networking sites concerning health care is that they do not illustrate or define completely the integrated solution of the problem stated. The user needs to navigate through multiple pages in order to arrive at a conclusion, which often becomes cumbersome at times. Hence, the goal of this research brief is to take a snapshot of current and emerging uses of social media and take a step toward the development of a crawling algorithm based on ontology to measure the impact of social media on healthcare. In this research I plan to analyze current and emerging uses of social media in the field of healthcare. Based on the observed trends I tend to propose a mining algorithm that would link the trends and suggest for best possible options to end-user based on geographical location. The algorithm is being proposed with the intent of crawling across social sites like Twitter, Facebook, Orkut, LinkedIn, etc. based on self-defined ontology of key medical or healthcare associated terms observed.

Chapter 1

INTRODUCTION

Discussions around personal health, and by extension healthcare, are no longer private issues. Citizens today are taking keen interest in their health and prefer sharing health related information with their peers. With governments across the world re-looking at their healthcare systems and taking measures in a bid to extend healthcare benefits to as many citizens as possible, public health and associated policy remains a key topic of discussion.[1]

In the healthcare industry, face-to-face interaction has traditionally been the primary medium for information exchange. The entire communication process has always revolved around a real-time physical setting in which patients met with physicians. However, newer technologies such as the internet and social media websites have enabled healthcare providers to reach their patients in a virtual environment, revolutionizing the way patient care is being provided.

An increasing number of people now use the web to communicate and are more likely to use social media websites and portals to procure healthcare-related information. A May 2011 ***Pew Research Centre's*** study showed that, of 3,001 US adults surveyed, approximately 80 percent of the internet users use online media for healthcare information. The same study also revealed that 15 percent of US adults use their cell phone to access healthcare information. A cross developing economies as well, using the internet to access healthcare information is fairly popular. An online survey conducted by Max Bupa Health Insurance in 2010 revealed that of 1,004 Indians surveyed, 39 percent use the internet for general healthcare information.[2]

The interaction of users with their physicians using social media are helping patients shift from the regular in-person patient–doctor interaction to a virtual environment where e-doctors provide solutions at lower costs. Technology has also facilitated the transformation of one-to-one communication into a one-to-many setup, in which a doctor reaches out to thousands of online users in seconds through tweets or blog posts. Interactive websites such as ***Hello Health*** and ***PatientsLikeMe*** enable users to easily connect with online doctors and patients to get instant information regarding various ailments.[2]

A recently concluded consumer survey in the US says:[1]

- 60 Million Americans exchanged their medical experiences online with each other last year.
- Almost 72 % of patients searched for online information before or after a doctor visit.

- 890 hospitals in the US used social media to engage with their patients.

Other key healthcare trends for 2010 include:[1]

- 73% of US consumers consider being physically fit important to being 'well', with 74% including 'feeling good about themselves'.
- An estimated 500 million people worldwide are expected to be using mobile healthcare applications by 2015.
- There were nearly 17,000 health apps available in major app stores in November 2010, with 57% of them being aimed at consumers rather than healthcare professionals.
- The heaviest use of health or medical related apps is by young adults: about 15% of those aged 18 to 29 have such apps, compared to 8% of users aged 30 to 49.

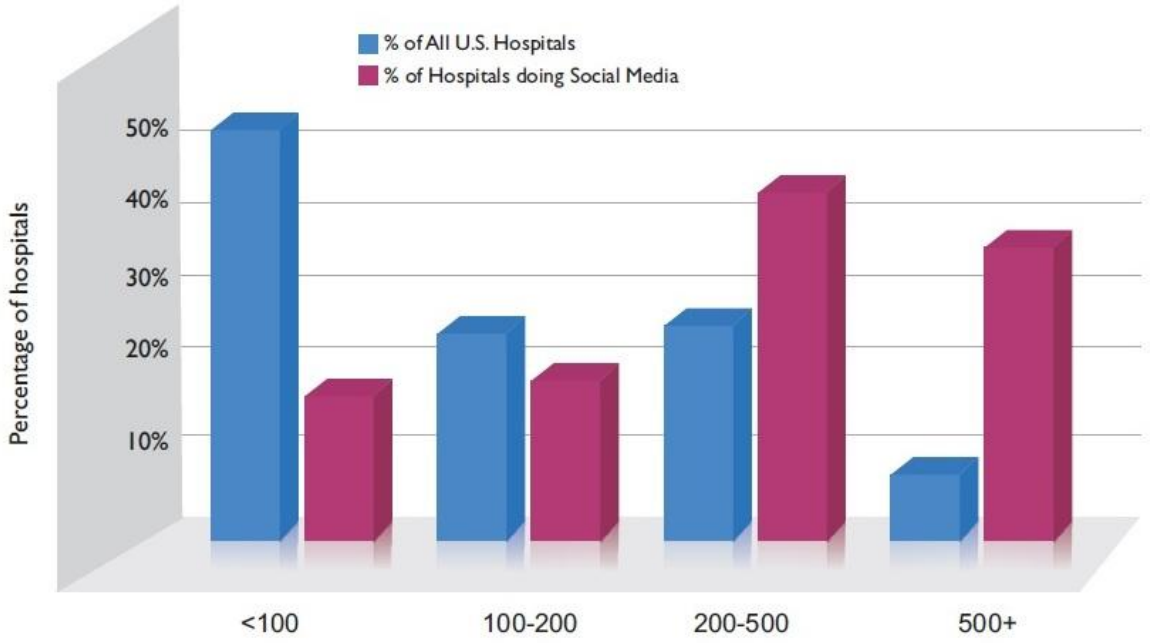


Figure 1- Depicts the trend of hospitals in USA w.r.t hospital care and impact of social media (based on Hospital Bed Count)

The figure clearly shows that larger hospitals are early adopters of Social Media. Hospitals with a bigger patient base have clearly more scope to engage in more user generated content. Also most big hospitals such as Mayo Clinic, who for the last 100 Years have relied on word of mouth to propagate their services, are rapidly adopting social media and have dedicated personnel who are engaging patients like never before.

Most of us probably think about Facebook and Twitter first when we hear the phrase social media. However when we consider the impact of social media on healthcare, we need to consider the powerful combination effect of these channels with:

- Blogs – most often by physicians, sometimes by patient advocates.
- Affinity group sites (condition-specific or role-specific).
- Reference sites (dot-coms, crowd sourced sites, patient experience rating sites).

A source, same as above also indicates that there are a total of 2,337 hospital social networking sites in the United States today. State wise, New York is the clear leader with almost 253 sites across various hospitals. California is second at 144. Channel wise, both Facebook and Twitter are equally popular, with YouTube and LinkedIn coming second. Blogs are a distant third.[1]

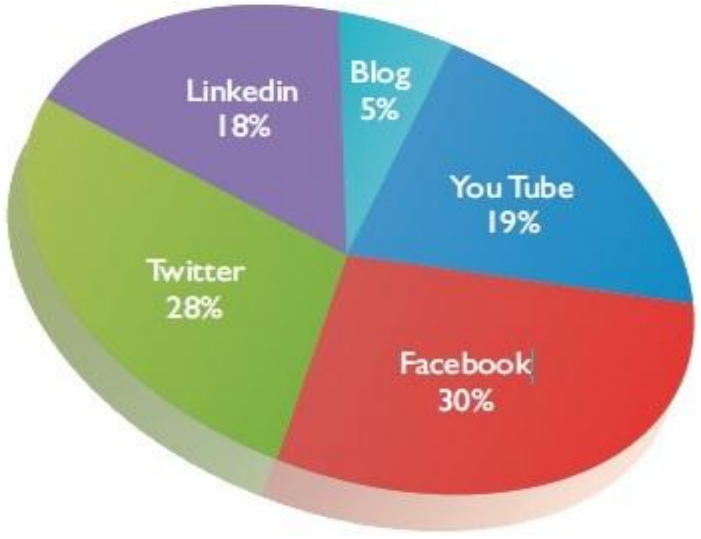


Figure 2- Distribution of healthcare related information across social networking sites

1.1 Points to ponder:

- During the ***H1N1 outbreak*** of 2009, the ***Centres for Disease Control and Prevention (CDC)*** turned to social networking site ***Twitter*** to communicate with clinicians across the US. The CDC employs a Twitter feed for emergency information (more than 1.2 million followers) and also for flu information (46,000 followers).[3]
- Patient Affinity Groups, form a leveraged crowd-sourcing opportunity for consumers to make informed choices about treatment and to learn from others who have been diagnosed with a

condition. Further, when they achieve critical mass (enough participants) they collect self-reported patient data about what works and what doesn't to form a growing self reported comparative effectiveness study. This is in fact the business model for several web companies today who manage these sites.[3]

- A social networking Web site, ***Facebook***, was used to deliver long-term sport psychology consultation services to student-athletes (i.e., soccer players) in 30- to 60-min weekly sessions. Additional short-term team building, group cohesion, communication, anger management, injury rehabilitation, mental toughness, commitment, and leadership workshops were provided. Cohesion and overall relationships between both the student-athletes and the sport psychology consultants benefited from this process. Social networking Web sites offer a practical way of providing sport psychology consulting services that does not require use of major resources.[4]
- Many clinics have been using the social networking sites for communicating with their patients, one such example is Society for Assisted Reproductive Technology (SART).[5]
- The UCLA HOPE (Harnessing Online Peer Education) study seeks to determine the acceptability, feasibility, and preliminary effectiveness of using online social networks (Facebook.com) to scale the Community Popular Opinion Leader (C-POL) model to increase HIV prevention among high-risk populations, with a focus on African American and Latino men who have sex with men (MSM). Participants join an online Facebook group related to HIV prevention and interact with community peer leaders trained in the fundamentals of HIV prevention and behaviour change over a 12-week period. The data collected from this study are compared to results from previous studies of community-based testing to preliminarily determine whether social networks can be used to scale community-based HIV prevention methods.[6]
- People with a regular health care provider, chronic disease, and those in younger age groups are more likely to consult online rankings and reviews and use SNS for health-related activities.[7]
- According to ***Phil Baumann,*** whose currently a member of Advisory Board at ***Mayo Clinic Center for Social Media*** reports that there are 140 health care uses for ***Twitter(***some of them are : disaster alerting and response, risk management communication, weight management and support, physician opinion sharing, tracking patient trends etc.).The healthcare picture would be totally different in future from what it seems today.[8]
- There's a web-site called ***twitterdoctors.net***, every hour, they provide an updated list of the most influential physicians on Twitter. The number of followers range from thousands to millions, and number of tweets are in the thousands to tens of thousands.[9]

- ***www.patientslikeme.com*** is a great example of a social network tool in this category.[10]
- Patient Monitoring takes advantage of the latest inventions and IT gizmos to help each of us stay on track in our lifestyle choices.[10]
- The recent use of networking sites aims to improve education, provide a forum to discuss relevant medical topics, and allow for improved patient care. The use of social media, with the understanding of its limitations, may help promote patient happiness and safety and serve as an educational platform.[11]

Table 1 - Online healthcare portals with different business models: [3]

Portal	Year of Introduction	Business Model	Revenue Model	Key differentiating factor
Hello Health(US)	2008	• Facilitates online communication between patients and physicians. • Provides the following practice solutions to healthcare providers: —***New direct-pay practice*****:** Hello Health optimizes the workflow of its clients by managing their documentation and patient profiles online. ***Addition of new cash-pay services to the existing insurance-based practice***: Through this, Hello Health allows its clients to add cash-pay services such as secure email and video visits to their existing insurance-based practice	Earns revenue by providing different membership plans to patients and practice solutions to healthcare providers	Easy-to-use online communication portal.

Healthcare Magic (India)	2008	• Helps patients find doctors by specialty in India, the UK and the US. • Facilitates online communication between patients and doctors. • Provides information on diseases and conditions	• Funded by sponsorships and advertisements. • Sells different membership plans to patients for interaction with doctors.	Provides personalized diet and fitness plan services.
Organizedw-isdom(US)	2007	• Provides healthcare information. • Collects and provides information shared by healthcare experts to users.	Funded by advertisements and sponsorships.	Free alerts are provided to users.
Sermo(US)	2006	• Facilitates physician–physician collaboration. • Provides healthcare information through dedicated blogs	Generates revenue from healthcare institutions, government agencies and financial services firms that purchase its products to access the elite group of practitioners.	Inter-doctor collaboration. Practicing physicians can join the portal for free.
Patients LikeMe (US)	2004	• Facilitates patient–patient interaction. • Provides healthcare information through research reports.	Sells patient experiences to companies that develop or sell medical products such as devices, equipment, drugs, medical services and insurance.	Dedicated focus on patients looking for other patients with similar ailments.

NHS Direct(UK)	1998	• Provides healthcare advice and information through telephone and online portal service to patients across the UK. • Helps patients find different healthcare providers operating in their region.	Earns revenues by providing patient care activities.	Provides service application for IPhones and Android mobiles.
RateMyMD(Canada)	NA	• Allows users to rate their doctors, based upon their experiences.	Funded by advertisements and sponsorships.	Provides ratings and feedback from user experience in terms of punctuality, knowledge, helpfulness and overall quality.

1.2 Why Social Media in Healthcare?

With the increased accessibility of social media globally, it is now being used as a tool to foster serious discussion on healthcare issues. Apart from being an always-on support platform for patients in need of advice and treatment, social media also offers significant advantages to healthcare service providers.[3]

Benefits for provider organizations:

- ***Reduced costs*** – Social networking platforms offer huge cost benefits to healthcare providers, as the two-way online communication helps minimize the costs associated with paper, telephone calls and other overhead components. The Department of Veterans Affairs (VA), a US-based military veteran benefit system, implemented VistA, an online healthcare system that enables doctors to interact with each other on a regular basis. According to a 2010 report by Healthcare Performance Management (HPM) Institute, a US-based research and education organization, the online communication system helped VA reduce its per-patient healthcare costs by 30 percent.[3]

- ***Improved physician–physician collaboration*** – Social media provides an ideal platform for healthcare professionals to share ideas, experiences and medical journal articles with each other. Websites such as Sermo and Ozmosis enable doctors to collaborate and share their favourite articles and research using social bookmarking functions, postings and voting tools. This helps ensure better patient outcomes, as doctors can conveniently build their knowledge in real time and facilitate faster adoption of best clinical practices.[3]

- ***Wider reach with no geographic boundaries*** – Social media has the capability to reach a large audience within a few seconds. In case of critical emergencies such as natural disasters and disease outbreaks, social media has proved to be an effective tool. During the devastating earthquake and tsunami that hit Japan in March 2011, Twitter acted as a lifesaver, as doctors could update chronically ill patients about various treatment locations. Also, during the 2009 H1N1 flu outbreak, the Centers for Disease Control and Prevention (CDC) used social media platforms to update users about the disease and its possible preventions. Its Facebook page was followed by over 50,000 users, and H1N1 videos on YouTube were viewed over 2.6 million times.[3]

- ***Launch of new services*** – Certain healthcare providers prefer using social media tools such as blogs to introduce new services. This helps the provider achieve better publicity than it can through a regular press release or other media channels. Users can easily comment and share their feedback on the new service. The Holy Cross Hospital in Florida has been regularly using social media to pass information on new therapeutic methods on its different therapeutic-specific blogs.[3]

- ***Talent search*** – As the healthcare industry requires specialized skills and expert knowledge, healthcare providers are moving beyond the traditional methods to search for workers with the required skill sets. Healthcare providers such as the Mayo Clinic and HCA Group both use Twitter to advertise new openings in their hospitals. Healthcare professionals are visiting social media platforms and professional portals such as LinkedIn to look for new opportunities. In 2010, AMN Healthcare conducted a study of more than 1,200 healthcare professionals, including doctors, pharmacists, nurses and allied professionals across multiple disciplines. The study revealed that around 20 percent of the respondents have used at least one social media website for their job search. As information can be disseminated rapidly through social media, healthcare providers are likely to leverage this further in their talent search programs.[3]

Benefits for Online Patients:

- ***Reduced costs*** – Social media initiatives help patients get timely solutions to their problems at lower costs. Instead of paying huge amounts of money to visit a doctor, a patient can easily register with a healthcare portal and get on-demand healthcare services at a fraction of the cost of a physical visit. A regular online visit to a doctor in the US is estimated to cost around US$30 per visit, much

less than the US$75–100 fee for an in-office visit.27 According to a US-based online healthcare service provider *Teladoc*, it provides 62 percent cost savings on visits to a regular primary care physician (PCP), 68 percent on visits to a specialist, 75 percent on visits to an urgent care unit and 95 percent on visits to an emergency room.[3]

- ***Improved self monitoring*** – Social media websites and healthcare portals such as WebMD provide a range of healthcare tools and applications that help the users to better understand their healthcare needs. WebMD tools like *Fit-O-Meter*, *Food and Fitness Planner* and *Personal Diet Evaluator* help the users conveniently monitor their daily healthcare and exercise requirements.[3]

- ***Interaction with other patients to make informed choices*** – Healthcare portals such as PatientsLikeMe and iMedix help patients with similar medical conditions to connect with each other, discuss their conditions and make informed choices regarding doctors and medical treatments. According to a June 2010 survey of 1,323 *PatientsLikeMe* members, 12 percent of patients changed their physicians as a result of interaction with other patients. The survey also highlighted that 22 percent of mood disorder patients required less in-patient care, subsequent to the regular interactions they had on the website.[3]

- ***Readily available healthcare provider information*** – Several social media websites and healthcare portals help patients looking for a hospital or a healthcare facility in a particular locality. Websites such as *ZocDoc* and *WebMD* allow users to select doctors as per their specifications, and also help in booking appointments. At the same time, there are several other portals such as *RateMyMD* and *DoctorScorecard* that allow users to rate their doctors, based upon their experiences.[3]

1.3 Future Implications of Social Media on Healthcare:

While it is important to get feedback from customers, organizations must know how to incorporate this feedback in a meaningful way in their business and brand strategies. In keeping with that, one can be expected to see the impact of social media in the following areas:

- ***Brand monitoring and management:*** Social Media in healthcare could become the number one source of brand monitoring. Currently not too many Healthcare firms are doing that. But organizations outside the healthcare industry, such as Dell, have built their reputation by monitoring their brand on social media. The company, which has been on social media since 2006,

mines data from all social networking sites and keeps track of what customers are talking about when it comes to their brand. Complaints are responded to, new product ideas are encouraged and orders are booked through Twitter (In 2009 the company generated business of $3 million through Twitter.) The company has also created a platform to syndicate content and aggregate content so as to monitor the brand. In the healthcare industry, the need for brand monitoring is not only to respond to a problem or complaint but also to measure marketing effectiveness and create new channels for visibility and eventually sales. For instance, with the coming up of the health insurance exchanges, individuals can buy health insurance on their own, without depending on their employers. This will open up a big channel for health plans across the US to effectively leverage social media and reach out to prospective customers with real time updates on aspects such as plans with the lowest premium for the day/ month/ quarter or lowest co-pays. Further they could also have special schemes for the holiday season or discounts on specific schemes, thus turning the social media channel into a revenue generating channel. Hospitals on the other hand should be able to offer services like online consultation, availability of appointments/ doctors, special discounts on the pharmacy products and booking operation theatres for minor procedures via social media like twitter. Blogs can be used for fostering research and collaboration across hospitals.[3]

- ***Faster industry cycles:*** In the last few decades, we have seen how the coming of age of computers helped reduce industry cycles and helped products go to market faster. The internet revolution resurrected the services business. In a similar way social media is today influencing industry cycles by:
  1. Helping bring out products faster – by enabling co-creation and collaboration.
  2. Customizing products to suit individuals – by providing data on individual preferences.

  The healthcare industry could benefit from both these aspects. In the area of health plans, the general move is towards offering customized plans to suit individuals and not offer them group plans. To do so, companies typically invest in third party research to obtain certain demographics and psychographic trends, which are then applied to tweak their existing plain vanilla insurance offerings. These tweaked plans are then tested in the market with focus groups to ascertain the potential for success. The entire process typically can take about 6- 12 months. This time frame can be reduced by half if companies start monitoring social media and engage in relevant conversations with consumers and prospects, seek opinions and incorporate these into customized plans. Using a content syndication and aggregation tool, not only would they be able to reduce dependence on

research firms, but also get closer to the consumer and launch more feasible products. In the medical devices industry, a lot of devices prescribed to patients for home use are not very user friendly. At some stage if patients are involved in a greater way in the manufacturing and design aspects of these devices, then the acceptance for these devices would be much higher.[3]

1.4 Brief Introduction of the Project:

Social media is a pervading business like never before. We all have used social networking sites like Facebook, Twitter, and LinkedIn for some or the other purposes, but a very small percentage of us would have used them for exchanging information regarding healthcare. Healthcare would not be the same paper-pen kind as what we see today, ten years thereafter.

Healthcare organizations and patients too need to understand how the chatter in the social media universe can be harnessed meaningfully. While some organizations have taken a lead in this area, many others are struggling to understand this new medium of opportunity. Developing a social media policy is no longer an option today, it is a necessity. If healthcare organizations do not take efforts in this direction, they run the risk of becoming stagnant and perhaps obsolete in the long run.

Healthcare nowadays, is no longer a private issue, instead individuals are taking keen interest in their health related data and prefer sharing the same with peers using social media. They tend to take a decision like hospital to visit, kind of treatment to be followed, particular doctor to consult, etc. based on the responses observed on social media.

1.4.1 Aim:

By looking at the current facts and figures of the growing users in healthcare based social media, the **Aim** of this research study is to analyze current and emerging uses of social media in the field of healthcare. Based on the observed trends we plan to develop a mining algorithm, that would link the trends and suggest for best possible options to end-user based on geographical location. The algorithm is being proposed with intent of crawling across social sites like Twitter, Facebook, Orkut, LinkedIn, etc. based on self-defined ontology of key medical or healthcare associated terms observed.

1.4.2 Different from existing approaches:

Healthcare based topics/comments are usually being discussed across different social networking sites like Facebook, twitter, LinkedIn, yahoo groups, Google groups, however they do not illustrate/define an integrated approach for the stated problem statement. We need to independently navigate through different sites to find an optimal solution, which often is a cumbersome process.

Henceforth, we propose for an integrated service that is based on self-designed ontology based on clinical/healthcare terms being used across the social sites.

1.4.3 Uniqueness of the approach:

Development of social media based clinical ontology is the key uniqueness of the proposed project. Another key feature being proposed in the study is a data mining algorithm that would recommend for best possible healthcare options to an end user based on given set of inputs at a particular geographical location.

1.4.4 Project Plan:

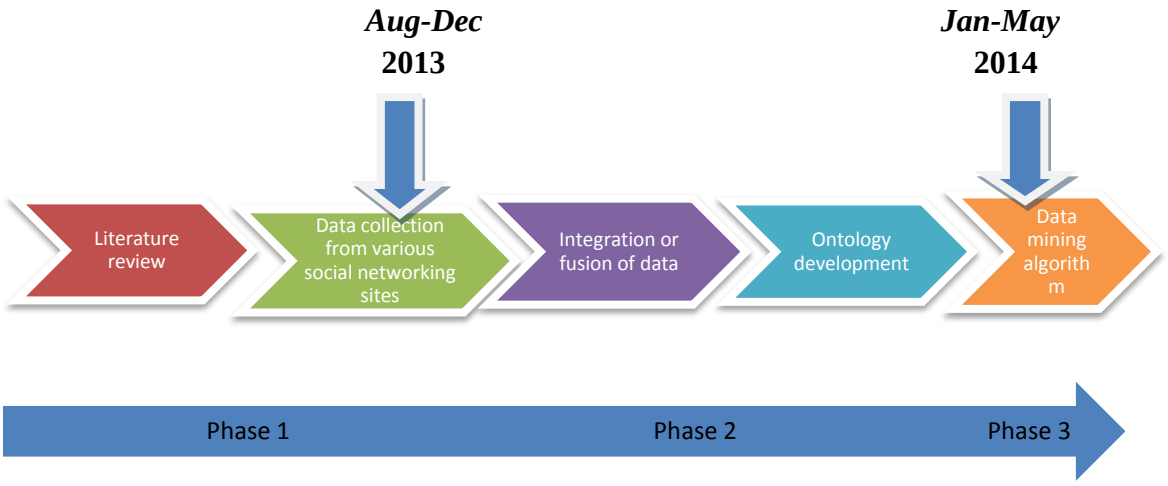


Figure 3 - Project Plan

Chapter-2

TOOLS AND TECHNIQUES

2.1 Pentaho

Pentaho Data Integration (PDI, also called Kettle) is the component of Pentaho responsible for the Extract, Transform and Load (ETL) processes. Though ETL tools are most frequently used in data warehouses environments, PDI can also be used for other purposes:

- Migrating data between applications or databases.
- Exporting data from databases to flat files.
- Loading data massively into databases.
- Data cleansing.
- Integrating applications.

PDI is easy to use. Every process is created with a graphical tool where you specify what to do without writing code to indicate how to do it; because of this, you could say that PDI is metadata oriented. PDI can be used as a standalone application, or it can be used as part of the larger Pentaho Suite. As an ETL tool, it is the most popular open source tool available. PDI supports a vast array of input and output formats, including text files, data sheets, and commercial and free database engines. Moreover, the transformation capabilities of PDI allow you to manipulate data with very few limitations.[12]

2.1.1 Kettle

Kettle is a free, open source (LGPL) ETL (Extraction, Transformation and Loading) tool. The product name should actually be spelled as K.E.T.T.L.E, which is a recursive acronym for "Kettle Extraction, Transport, Transformation and Loading Environment".

Kettle was first conceived about four years ago by Matt Casters, who needed a platform-independent ETL tool for his work as a BI Consultant. Matt's now working for Pentaho as Chief of Data Integration.

Being an ETL tool, Kettle is an environment that's designed to:

- Collect data from a variety of sources (extraction).
- Move and modify data (transport and transform) while cleansing, denormalizing, aggregating and enriching it in the process.

- Frequently (typically on a daily basis) store data (loading) in the final target destination, which is usually a large, dimensionally modelled database called a data warehouse

Although most of these concepts are equally applicable to almost any data importing or exporting processes, ETL is most frequently encountered in data warehousing environments.[12]

2.1.2 Kettle Architecture:

Kettle is built with the java programming language. It consists of four distinct applications:

Spoon-

It is a graphically oriented end-user tool to model the flow of data from input through transformation to output. One such model is also called a transformation.

Pan-

It is a command line tool that executes transformations modelled with Spoon.

Chef-

It is a graphically oriented end-user tool used to model jobs. Jobs consist of job entries such as transformations; FTP downloads etc. that are placed in a flow of control.

Kitchen-

It is a command line tool used to execute jobs created with Chef.

Model-driven

An interesting feature of Kettle is that it is model-driven. Both Spoon and Chef offer a graphical user interface to define the ETL processes on a high level. Typically, this involves no actual programming at all - rather, it's a purely declarative task which results in a model.

The command line tools Pan and Kitchen (or rather the underlying API) know how to read and interpret the models created by Spoon and Chef respectively. These tools actually execute the implied ETL processes. This is done all in one go: there is no intermediate code generation or compilation involved.[12]

Repository-Based

Models can be saved to file in a particular XML format, or they can be stored into a relational database: a repository. Using a repository can be a major advantage, especially when handling many models. Because the models are stored in a structured manner, arbitrary queries can be written against the repository. The repository may also be used to store the logs that are generated when executing

transformations and jobs. Certain environments, such as banks, require that every manipulation that is performed with financial data be stored for longer periods of time for auditing purposes. The repository sure seems to be the place to do that, at least, as far as the ETL process is concerned.[12]

Pentaho Usage:

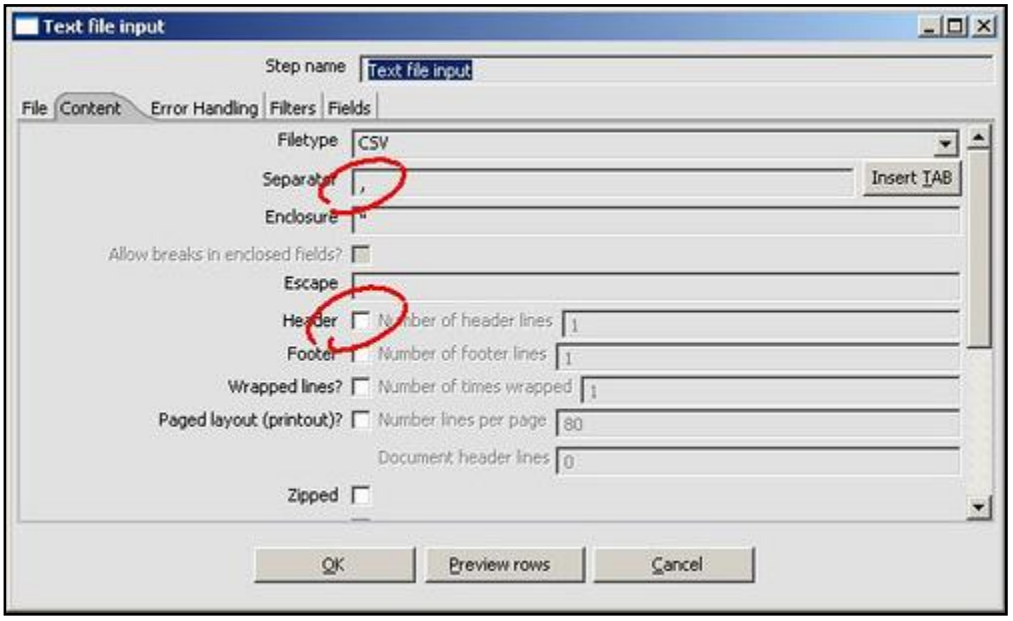


Figure 4 - Loading an Input file dialog box

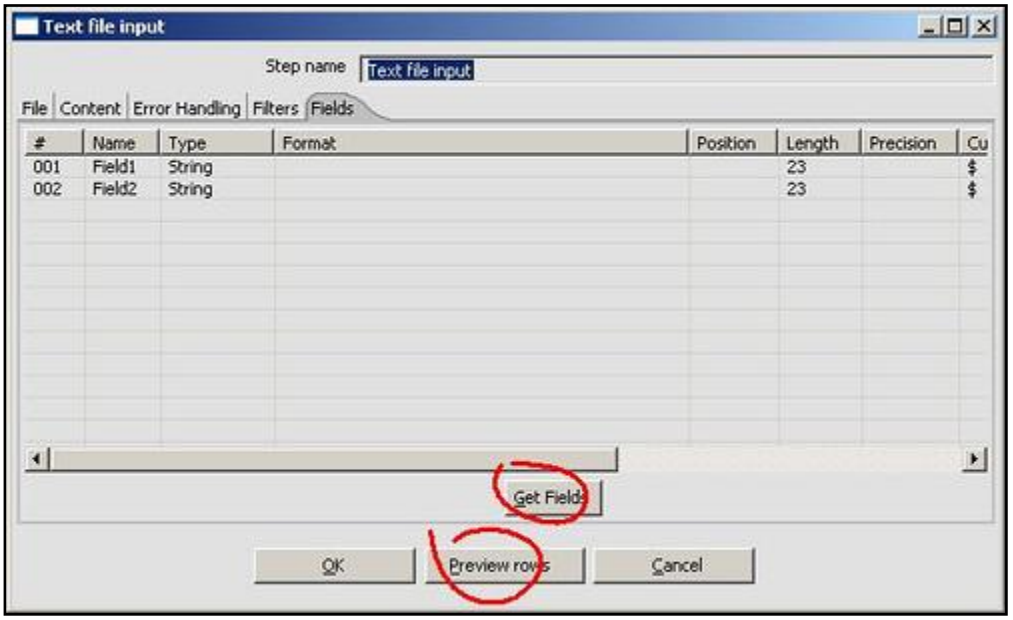


Figure 5 - Previewing the list of attributes

Figure 6 - Preview of data will look like in a table

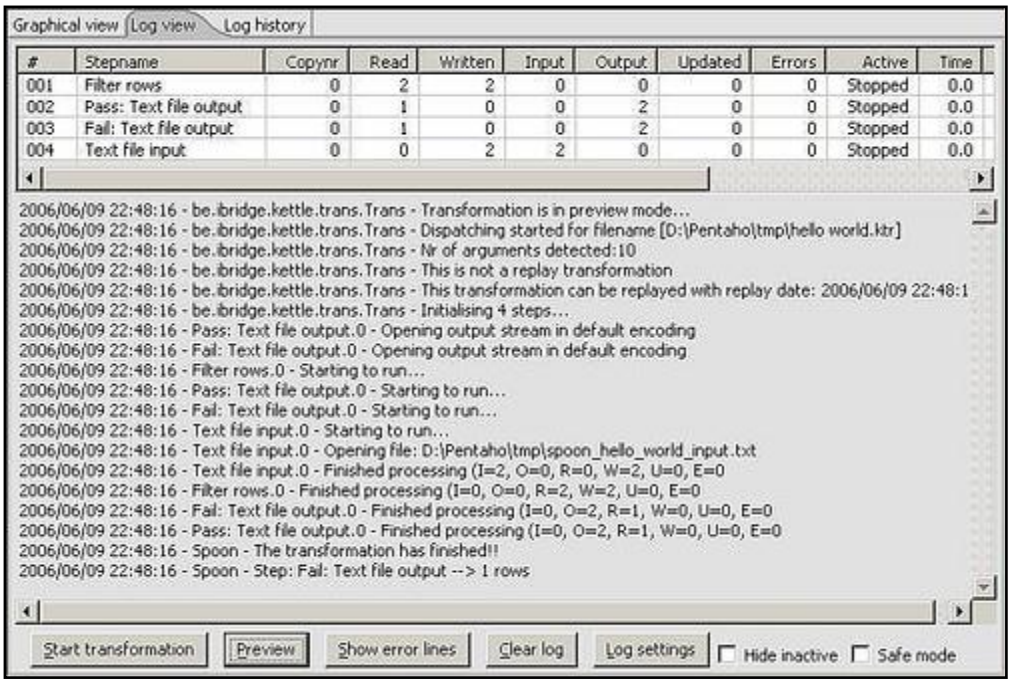


Figure 7 - List of errors along with the description

2.2 MySQL (RDBMS Package)

Modern day web sites seem to be relying more and more on complex database systems. These systems store all of their critical data, and allow for easy maintenance.

MySQL is a powerful Relational Database Management System (RDBMS), which uses the principles of database and data manipulation using Structured Query Language (SQL) statements. SQL is a database language that is used to retrieve, insert, delete and update stored data and its standardization makes it quite easy to store, update and access data. One of the most powerful SQL servers out there is called MySQL and surprisingly enough, it's free. Some of the features of MySQL Include: Handles

large databases, in the area of 50,000,000+ records. No memory leaks. Tested with a commercial memory leakage detector (purify). It has a privilege and password system which is very flexible and secure, and which allows host-based verification. Passwords are secure since all password traffic when connecting to a server is encrypted.

2.3 HTML

HTML (Hypertext Markup Language) is used to create document on the World Wide Web. It is simply a collection of certain key words called 'Tags' that are helpful in writing the document to be displayed using a browser on Internet. It is a platform independent language that can be used on any platform such as Windows, Linux, Macintosh, and so on. To display a document in web it is essential to mark-up the different elements (headings, paragraphs, tables and so on) of the document with the HTML tags. To view a mark-up document, user has to open the document in a browser. A browser understands and interpret the HTML tags, identifies the structure of the document (which part are which) and makes decision about presentation (how the parts look) of the document. HTML also provides tags to make the document look attractive using graphics, font size and colours. User can make a link to the other document or the different section of the same document by creating Hypertext Links also known as Hyperlink.

HTML instructions divide the text of a document into blocks called elements. These can be divided into two broad categories -- those that define how the BODY of the document is to be displayed by the browser and those that define information `about' the document, such as the title or relationships to other documents.

HTML-CSS

CSS- CSS is the abbreviation for Cascading Style Sheet. A style sheet simply holds a collection of rules that we define to enable us to manipulate our web pages. CSS is a style language that defines layout of HTML documents. For example, CSS covers fonts, colours, margins, lines, height, width, background images, advanced positions and many other things. HTML can be (mis-)used to add layout to websites. But CSS offers more options and is more accurate and sophisticated. CSS is supported by all browsers today.[13]

What is the difference between CSS and HTML?

HTML is used to structure content. CSS is used for formatting structured content. The language HTML was only used to add structure to text. CSS are a way to control the look and feel of your HTML documents in an organized and efficient manner. CSS was a revolution in the world of web design. The concrete benefits of CSS include:

- Control layout of many documents from one single style sheet;
- More precise control of layout;
- Apply different layout to different media-types (screen, print, etc.);
- Numerous advanced and sophisticated techniques.

2.4 PERL and CGI

2.4.1 Definition

CGI is the Common Gateway Interface, a standard for programs to interface with information servers such as HTTP (web) servers. CGI allows the HTTP server to run an executable program or script in response to a user request, and generate output on the fly. This helps in creating dynamic and interactive web pages. CGI programs can be written in any language. Perl is a very common language for CGI programming as it is largely platform independent and the language's features make it very easy to write powerful applications. It is important to remember that CGI is not a language in itself. CGI is merely a type of program which can be written in any language.

2.4.2 How CGI works?

When we click a hyper link to browse a particular web page or URL, the following steps are noticed:

- Browser contacts the HTTP web server and demand for the URL ie. filename.
- Web Server will parse the URL and will look for the filename in if it finds that file then sends back to the browser otherwise sends an error message indicating that you have requested a wrong file.
- Web browser takes response from web server and displays either the received file or error message.

However, it is possible to set up the HTTP server so that whenever a file in a certain directory is requested that file is not sent back; instead it is executed as a program, and whatever that program outputs is sent back for the browser to display. This function is called the Common Gateway Interface or CGI and the programs are called CGI scripts. [14]

2.4.3 Requirements to run Perl CGI programs

- A web server
- Web server configuration which gives permission to run CGI
- A Perl interpreter
- Appropriate Perl modules, such as CGI.pm
- A shell account is extremely useful but not essential

Most of the above requirements will need one's system administrator or ISP to set accordingly. Some will be wary of allowing users to run CGI programs, and may require obeying certain security regulations or paying extra for the privilege.

Chapter-3

METHODOLOGY AND CONCERNING LITERATURE

3.1 Literature

3.1.1 Introduction to Database

During the course of a day we might use a telephone directory, a dictionary, an encyclopaedia, an airline flight guide, a bibliography, Wikipedia, or a Yahoo or Google index. We use databases to store all kinds of knowledge that we retrieve on a regular basis. In fact, we use them so much we don't even think about the fact that we are using databases.

A ***database*** is an organized collection of facts and information. Databases usually contain text and numbers, and frequently they hold still images, sounds and video or film clips. A database permits its user to extract a specific group of disparate facts from within a collection of facts.

Formally, "database" refers to the data themselves and supporting data structures. Databases are created to operate large quantities of information by inputting, storing, retrieving and managing that information. Databases are set up so that one set of software programs provides all users with access to all the data.

A "database management system" (DBMS) is a suite of computer software providing the interface between users and a database or databases. Because they are so closely related, the term "database" when used casually often refers to both a DBMS and the data it manipulates.

Outside the world of professional information technology, the term database is sometimes used casually to refer to any collection of data (perhaps a spreadsheet, maybe even a card index).

The interactions catered for by most existing DBMSs fall into four main groups:

- Data definition – Defining new data structures for a database, removing data structures from the database, modifying the structure of existing data.
- Update – Inserting, modifying, and deleting data.
- Retrieval – Obtaining information either for end-user queries and reports or for processing by applications.

- Administration – Registering and monitoring users, enforcing data security, monitoring performance, maintaining data integrity, dealing with concurrency control, and recovering information if the system fails.

A DBMS is responsible for maintaining the integrity and security of stored data, and for recovering information if the system fails.

Both a database and its DBMS conform to the principles of a particular database model. "Database system" refers collectively to the database model, database management system, and database. Physically, database servers are dedicated computers that hold the actual databases and run only the DBMS and related software. Database servers are usually multiprocessor computers, with generous memory and RAID disk arrays used for stable storage. RAID is used for recovery of data if any of the disks fail. Hardware database accelerators, connected to one or more servers via a high-speed channel, are also used in large volume transaction processing environments. DBMSs are found at the heart of most database applications. DBMSs may be built around custom multitasking kernel with built-in networking support, but modern DBMSs typically rely on a standard operating system to provide these functions. Since DBMSs comprise a significant economical market, computer and storage vendors often take into account DBMS requirements in their own development plans.

Databases and DBMSs can be categorized according to the database models that they support (such as relational or XML), the types of computer they run on (from a server cluster to a mobile phone), the query languages used to access the database (such as SQL or XQuery), and their internal engineering, which affects performance, scalability, resilience, and security.

3.1.2 What is 'Ontology'?

It is becoming an impossible task for managing research data in the field of biomedical and clinical informatics. These huge data sets cannot manually be analyzed, interpreted or processed to acquire inferred knowledge efficiently. We need intelligent agents or computer systems to help us in doing these tasks and hence it becomes mandatory to represent medical knowledge in computer process able format. Semantic technology and ***ontology*** can be used to partial solve the data management problem in medical informatics. Semantic knowledge representation allows the intelligent agents or computers to interpret the data and acquire inferred knowledge. Hence, ***ontology design is an important aspect of medical informatics***, and reusability is a key issue that is determined by the level of compatibility among ontology concepts and among the theories of the biomedical domain.

Ontology is "***hierarchal structuring of knowledge about concepts by sub-classing them according to their properties and qualities***". It can also be defined as "a ***declarative model of a domain that defines and represents the concepts existing in that domain, their attributes and the relationships between them***". Ontology gives the description of concepts and the relations that can exist between them. The concept is very important for data sharing and knowledge representation.[15]

3.1.2.1 Classification

Ontology can be classified according to level of detailed knowledge they provide. ***Upper Ontologies*** provides very generic knowledge with low domain specific knowledge. For example, ***Disease ontology*** is upper ontology compatible for any biomedical domain. ***General ontologies*** represent knowledge at an intermediate level of detail independently of a specific task. ***Domain ontologies*** represent knowledge about a particular part of the world, such as medicine, and should reflect the underlying reality through a theory of the domain represented. For example, Gene Ontology, Finally, ontologies designed for specific tasks are called ***application ontologies.[15]***

3.1.2.2 Description

In ontology, concepts of the domain are represented by "classes". The features and attributes of the concept are described by "properties or slots". Together with "instances" which are individual of a class it constitutes the knowledge base of the domain. Classes are the main focus in ontology. Classes can be sub-classed to describe more specific features of a class. For example, if we define a class *Wines*, it includes all the wine classes in the wine domain. The wine class can be sub-classed to specify more specific wines like *Red Wine, White Wine* etc. Instances are individuals related to a same class. For example, *Australian Yellow Tail* is an individual for Wine class. Slots or Properties can be created to describe properties of a class or instance. For example, we can define a property named "*Has_Color*" which holds the color of a particular wine class or instance.

The figure shows the summary of the above discussion. Therefore in particular describing a domain in ontology includes:

- Defining the concepts of the domain as classes.
- Defining the individuals of the class as instances.
- Defining the attributes of the individuals as properties.
- Filling the properties values for the instances.

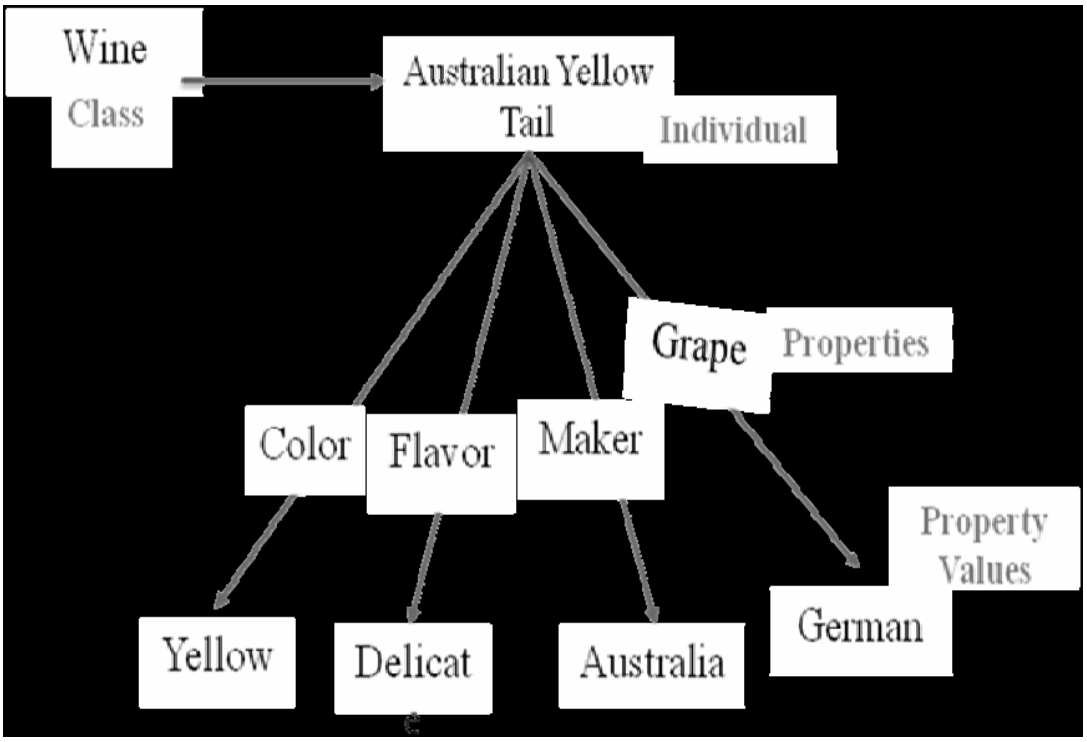


Figure 8 - Describing Australian Yellow Tail Wine instance and filling property values

3.1.2.3 Advantages

Ontologies are developed and defined to share the knowledge among the researchers working on the same domain. The main reasons for developing ontology are:

- Sharing the knowledge in the same domain is one of the common goal for which ontologies are developed. For example, many websites provide medical information about various concepts in the medical domain. If they use the same medical terms for describing the information, the data can be integrated easily from different sources by the computer agents and solve the user queries.
- To reuse the already build ontology is the recent tide in the ontology research. For example, the notation of capturing time has become important main. The notation of time includes time intervals, time of event occurrence, interval between events etc. If a group of researchers define an ontology which includes all these features, other research groups can simply import the existing ontology for serving the purpose.
- To separate domain knowledge from operation knowledge. For example, we can describe a task of configuring a product from its components according to a required specification and implement a

program that does this configuration independent of the products and components themselves. We can then develop ontology of PC-components and characteristics and apply the algorithm to configure made-to-order PCs. We can also use the same algorithm to configure elevators if we "feed" elevator component ontology to it. The fundamental rules that should be in mind when developing ontology should be:

- There is more than one way to define the same working domain. The best solution always depends on the application where the ontology is deployed.
- Development of ontology is iterative process.
- Before developing the ontology, the working domain range should be specified. This will give the scope of the ontology.

  Consider reusing the ontology. It's better to look into others work in the domain to see if we can directly use the ontology or refinement to the ontology can solve the problem.[15]

3.1.2.4 Present Biomedical Ontologies

For the past few years numerous ontologies have been developed in biomedical community with a soul aim, to represent biomedical terminology in a common vocabulary so that they can be shared and reused across various fields. These ontologies are developed by various medical centres, researchers, industries etc.

3.1.2.4.1 OpenCyc

OpenCyc is an Upper level ontology developed by *Cycorp* Inc. This project was started in 1984, has thousands of hand coded assertions that capture "common sense language". AI algorithms can perform human like reasoning on these assertions. OpenCyc is an upper level ontology, with 6,000 concepts and 60,000 assertions about those concepts.[15]

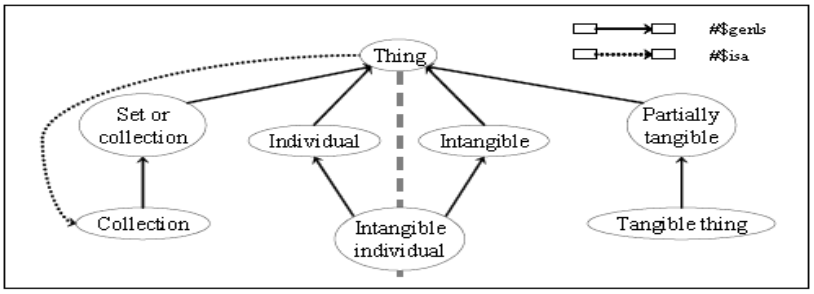


Figure 9 - OpenCyc top-level ontology

The figure shown above illustrates OpenCyc ontology with "Thing" as universal set of collections.

3.1.2.4.2 WordNet

WordNet is an electronic lexical database developed at Princeton University that serves as a resource for applications in natural language processing and information retrieval. The core structure in WordNet is a set of synonyms (Sysnet) that represents one underlying concept. Sysnet formation is based on synonymy (one meaning expressed by several words) and polysemy (one word having several distinct meanings). WordNet has been influenced by cognitive psychology as well as linguistics, and its hierarchies are not based on formal ontology theory. Many concepts that represent health disorders in medical terminologies, when present in WordNet, are categorized appropriately. For example, when keyword cancer has the following meaning in WordNet:

Cancer, malignant neoplastic disease: any malignant growth or tumor caused by abnormal and uncontrolled cell division; it may spread to other parts of the body through the lymphatic system or the blood stream

- Cancer, Crab: (astrology) a person who is born while the sun is in Cancer.
- Cancer: a small zodiacal constellation in the northern hemisphere; between Leo and Gemini.
- Cancer, Cancer the Crab, Crab: the fourth sign of the zodiac; the sun is in this sign from about June 21 to July 22.
- Cancer, genus Cancer: type genus of the family Cancridae.[15]

3.1.2.4.3 SNOMED- CT

SNOMED stands for **S**ystemized **N**omenclature **O**f **M**edicine **C**linical **T**erms. SNOMED-CT is the result of merging two ontologies: SNOMED-RT and Clinical Terms. SNOMED CT is the most comprehensive biomedical terminology recently developed in native description logic formalism.[15]

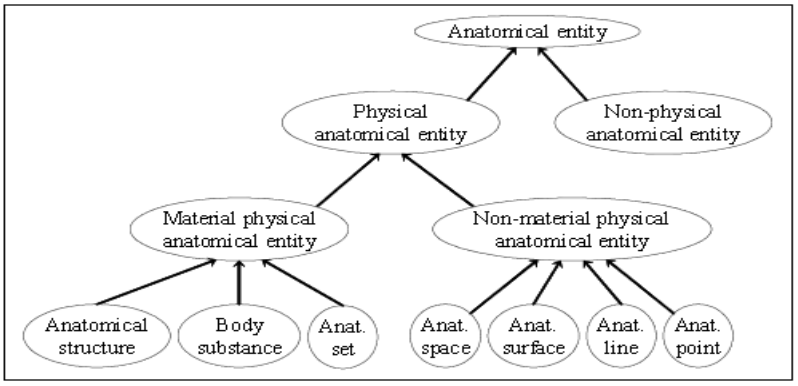


Figure 10 - SNOMED-CT ontology classification

3.2 Methodology

Social media isn't a fad; it's a fundamental shift in the way we communicate. It's a startling fact that there are around 800 million active users on Facebook. Out of which 180 million users login on a daily basis. It's also believed that if Facebook were a country, it would be third largest in the world. We no longer search news or products; instead they find us and that too on social media. If this medium is able to connect so many people of different age-groups, then why not make this medium a platform to exchange information regarding healthcare. This inspired me to work on this project.

3.2.1. Problem Analysis

I initiated the project by analyzing the current problem which prevails by scrutinizing the social networking sites concerning healthcare information.

Healthcare based topics/comments are usually being discussed across different social networking sites like Facebook, Twitter, LinkedIn, Yahoo groups, Google groups, however they do not illustrate/define an integrated approach for the stated problem. We need to navigate independently through different sites to find an optimal solution, which often is a cumbersome process. I also went through a lot of literature work for a deeper understanding on this issue.

3.2.2 Data Collection

Then comes the most important part of my research i.e. *Data collection* from various social networking sites. I couldn't find any relevant data from LinkedIn, Google groups and Yahoo groups; therefore I collected data from Facebook and Twitter only.

I first analysed what kind of data I require by looking into the networking sites having healthcare information. After doing this research, I then created a Performa in excel sheet which consisted of parameters like:

- Source
- Resource
- URL
- Type of information
- Image
- Location
- Description

Some snapshots of Performa are:

	A	B	C	D
1	SOURCE	RESOURCE	URL	TYPE OF INFORMATION
2	Facebook	Apollo Hospitals	https://www.facebook.com/TheApolloHospitals	awareness
3	Facebook	Apollo Hospitals	https://www.facebook.com/TheApolloHospitals	health tips
4	Facebook	Apollo Hospitals	https://www.facebook.com/TheApolloHospitals	promotional
5	Facebook	Apollo Hospitals	https://www.facebook.com/TheApolloHospitals	question n answer
6				
7	Facebook	Fortis Hospitals	https://www.facebook.com/fortishealth	recommendations
8	Facebook	Fortis Hospitals	https://www.facebook.com/fortishealth	many facts about alzheimer's disease
9	Facebook	Fortis Hospitals	https://www.facebook.com/fortishealth	advice tips by celebrities
10	Facebook	Fortis Hospitals	https://www.facebook.com/fortishealth	promotional
11				
12	Facebook	Lilavati Hospital	https://www.facebook.com/LilavatiHospitalResearchCentre	awards n honours
13	Facebook	Lilavati Hospital	https://www.facebook.com/LilavatiHospitalResearchCentre	health advice
14				
15	Facebook	CMC , Vellore	https://www.facebook.com/cmcvelloreindia	announcements/declarations
16	Facebook	CMC , Vellore	https://www.facebook.com/cmcvelloreindia	events organised
17	Facebook	CMC , Vellore	https://www.facebook.com/cmcvelloreindia	charity
18	Facebook	CMC , Vellore	https://www.facebook.com/cmcvelloreindia	awards n honours
19				
20	Facebook	Artemis Hospitals	https://www.facebook.com/ArtemisHospitals	events organised
21	Facebook	Artemis Hospitals	https://www.facebook.com/ArtemisHospitals	educational
22	Facebook	Artemis Hospitals	https://www.facebook.com/ArtemisHospitals	vacancy
23	Facebook	Artemis Hospitals	https://www.facebook.com/ArtemisHospitals	global concern
24	Facebook	Artemis Hospitals	https://www.facebook.com/ArtemisHospitals	various programmes
25	Facebook	Artemis Hospitals	https://www.facebook.com/ArtemisHospitals	offer training
26				

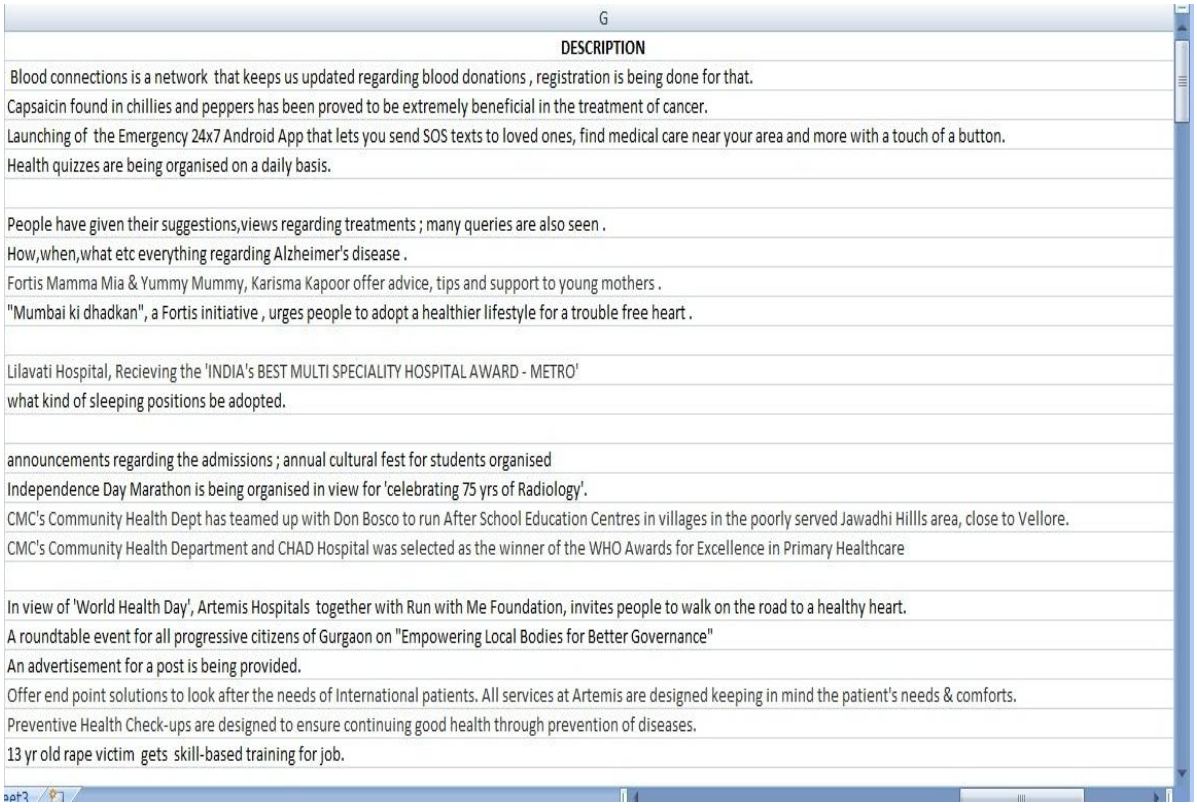
Sheet1 Sheet2 Sheet3

Figure 11 - Snapshot of Performa(1)

	E	F
1	IMAGE	LOCATION
2	https://m.ak.fbcdn.net/sphotos-h.ak/hphotos-ak-prn2/p480x480/1240624_566764023384710_1203879273_n.jpg	Across various cities in India
3	https://m.ak.fbcdn.net/sphotos-d.ak/hphotos-ak-frc3/p480x480/1235236_572110956183350_1592913626_n.jpg	Across various cities in India
4	https://m.ak.fbcdn.net/sphotos-a.ak/hphotos-ak-prn1/p480x480/994355_571626179565161_371583510_n.jpg	Across various cities in India
5	https://m.ak.fbcdn.net/sphotos-d.ak/hphotos-ak-frc1/p480x480/1185170_569747623086350_1684816815_n.jpg	Across various cities in India
6		
7	https://www.facebook.com/fortishealth	Mumbai
8	https://m.ak.fbcdn.net/sphotos-c.ak/hphotos-ak-prn2/p480x480/1239032_10151583002810951_1851261895_n.jpg	Mumbai
9	https://m.ak.fbcdn.net/sphotos-a.ak/hphotos-ak-ash3/p480x480/1234340_10151578221215951_1696036188_n.jpg	Mumbai
10	https://m.ak.fbcdn.net/sphotos-d.ak/hphotos-ak-ash4/1239990_10151576396530951_1465233009_n.jpg	Mumbai
11		
12	https://m.ak.fbcdn.net/sphotos-f.ak/hphotos-ak-frc1/p480x480/429765_366458443469436_785071126_n.jpg	Mumbai
13	https://m.ak.fbcdn.net/sphotos-c.ak/hphotos-ak-frc1/581685_428054473976499_1603430316_n.jpg	Mumbai
14		
15	https://m.ak.fbcdn.net/sphotos-h.ak/hphotos-ak-frc1/p480x480/1001920_606086119431920_971035643_n.jpg	Vellore
16	https://m.ak.fbcdn.net/sphotos-f.ak/hphotos-ak-ash4/s526x296/1003466_700017296682455_1267817092_n.jpg	Vellore
17	https://m.ak.fbcdn.net/sphotos-g.ak/hphotos-ak-ash4/p480x480/487528_549238188450047_1597090328_n.jpg	Vellore
18	https://m.ak.fbcdn.net/sphotos-g.ak/hphotos-ak-ash4/p480x480/482673_545034625537070_39316464_n.jpg	Vellore
19		
20	https://m.ak.fbcdn.net/sphotos-d.ak/hphotos-ak-prn1/1174556_572013816179157_839387774_n.jpg	Gurgaon
21	https://m.ak.fbcdn.net/sphotos-d.ak/hphotos-ak-ash3/p480x480/1175564_565853903461815_333703181_n.jpg	Gurgaon
22	https://m.ak.fbcdn.net/sphotos-b.ak/hphotos-ak-ash3/p480x480/542205_562885030425369_1019395756_n.jpg	Gurgaon
23	https://m.ak.fbcdn.net/sphotos-b.ak/hphotos-ak-ash4/1000058_539345232779349_668545422_n.jpg	Gurgaon
24	https://m.ak.fbcdn.net/sphotos-h.ak/hphotos-ak-frc3/1044388_539245786122627_1551105488_n.jpg	Gurgaon
25	https://m.ak.fbcdn.net/sphotos-a.ak/hphotos-ak-ash3/p480x480/941128_533456356701570_662966277_n.jpg	Gurgaon
26		

Sheet1 Sheet2 Sheet3

Figure 12 - Snapshot of Performa(2)



G
DESCRIPTION
Blood connections is a network that keeps us updated regarding blood donations , registration is being done for that.
Capsaicin found in chillies and peppers has been proved to be extremely beneficial in the treatment of cancer.
Launching of the Emergency 24x7 Android App that lets you send SOS texts to loved ones, find medical care near your area and more with a touch of a button.
Health quizzes are being organised on a daily basis.
People have given their suggestions,views regarding treatments ; many queries are also seen .
How,when,what etc everything regarding Alzheimer's disease .
Fortis Mamma Mia & Yummy Mummy, Karisma Kapoor offer advice, tips and support to young mothers .
"Mumbai ki dhadkan", a Fortis initiative , urges people to adopt a healthier lifestyle for a trouble free heart .
Lilavati Hospital, Recieving the 'INDIA's BEST MULTI SPECIALITY HOSPITAL AWARD - METRO'
what kind of sleeping positions be adopted.
announcements regarding the admissions ; annual cultural fest for students organised
Independence Day Marathon is being organised in view for 'celebrating 75 yrs of Radiology'.
CMC's Community Health Dept has teamed up with Don Bosco to run After School Education Centres in villages in the poorly served Jawadhi Hillls area, close to Vellore.
CMC's Community Health Department and CHAD Hospital was selected as the winner of the WHO Awards for Excellence in Primary Healthcare
In view of 'World Health Day', Artemis Hospitals together with Run with Me Foundation, invites people to walk on the road to a healthy heart.
A roundtable event for all progressive citizens of Gurgaon on "Empowering Local Bodies for Better Governance"
An advertisement for a post is being provided.
Offer end point solutions to look after the needs of International patients. All services at Artemis are designed keeping in mind the patient's needs & comforts.
Preventive Health Check-ups are designed to ensure continuing good health through prevention of diseases.
13 yr old rape victim gets skill-based training for job.

Figure 13 - Snapshot of Performa(3)

3.2.3 Creation of Database

The data was in the form of images, text, and other parameters, which were filled in the excel sheet Performa. I first converted the data in the excel sheet into comma separated file format (CSV). After which it was processed into the ‘staging database’.

The ***staging database*** is a separate data cache (storage area) that helps users in continuous access to application data. Their access continues even when data is being imported from the various external sources and prepared for loading. This minimizes the downtime that users experience during data loading or data refreshing. Here the data is dumped as it is, without any changes being made to it i.e. the data here is in its original form. I then extracted the data from the staging schema and then processed it, cleaned it so that it could further be used for analysis. The process of cleaning and transforming data is known as ETL, or Extraction, Transformation, and Loading.

I have used ***Pentaho 3.0 (Kettle)*** and ***MySQL*** to preprocess and store the data respectively. A staging database in MySQL by the name “***socialhealth***” was created which holds a table named “**test_db**”. The next step was to designs mappings (transformation) in Pentaho. Transformation applies a series of rules

or functions to the extracted data from the source to derive the data for loading into the end target. Some data sources required very little or even no manipulation of data.

I made a transformation by the name "***snh_data***" in which I processed data which was in the .csv format was loaded into the "***snh_table***" table in ***socialhealth*** database. The transformation looked like:

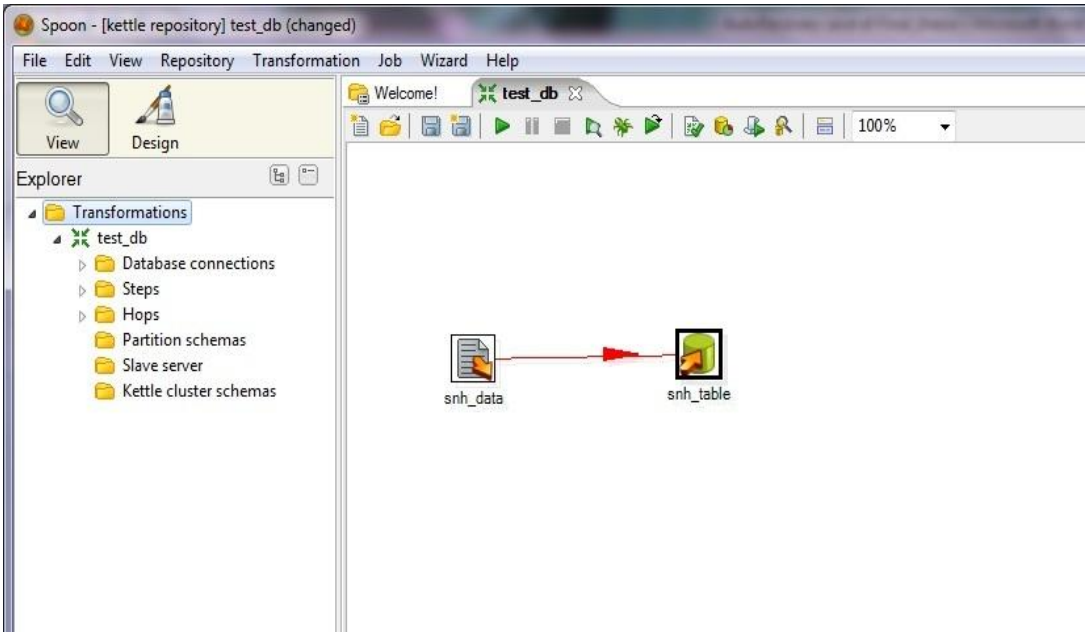


Figure 14 - Snapshot of a Transformation

The "snh_data" contains the csv format file where we mentioned various details such as the csv file name, name of this input etc.

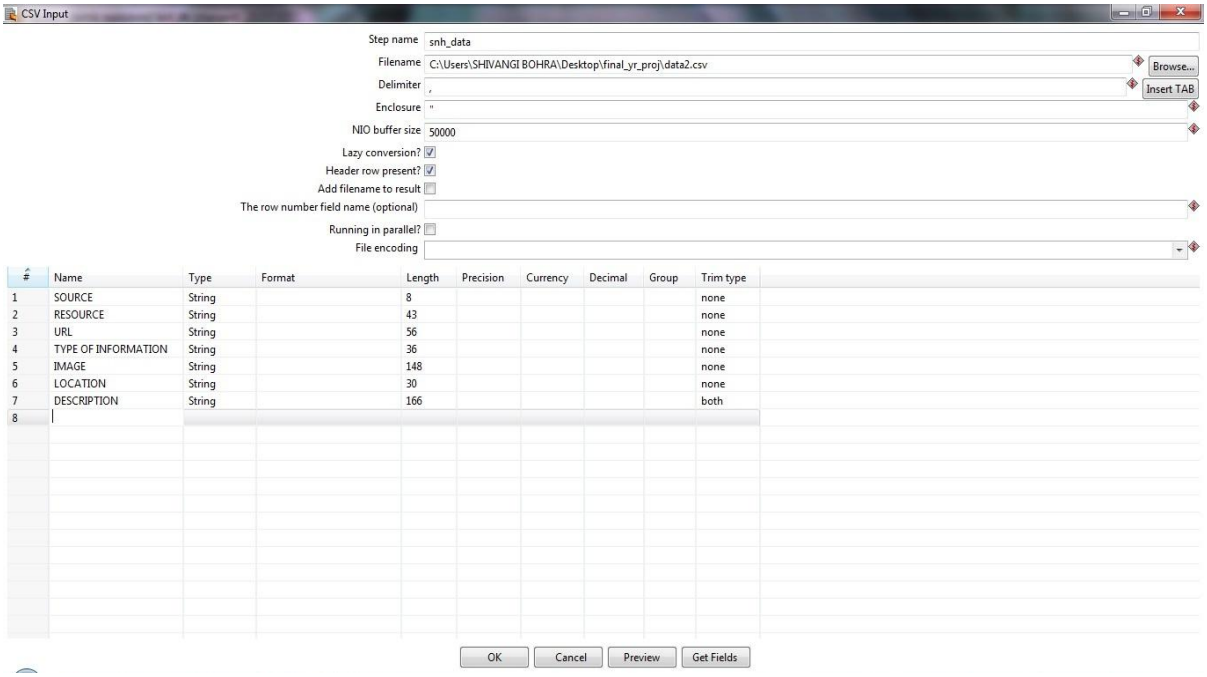


Figure 15 - Snapshot of an Input File

Spoon has the option of "Get Fields" where we can get all the attributes from the csv file by analyzing a few number of records from our data .It asks the user the number of sample it wants to be analyzed in order to get the attributes.

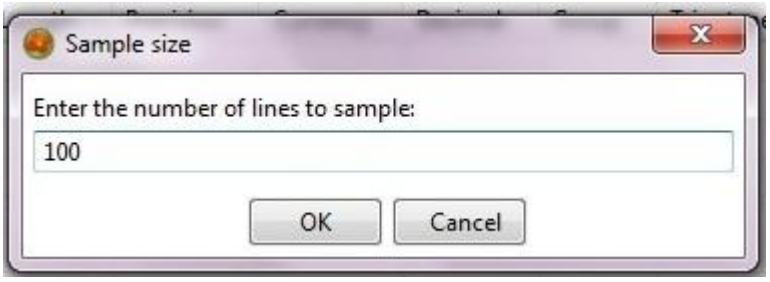


Figure 16 - No .of samples to be analyzed to get fields

It then shows you a window where all the possible attributes which the spoon has found are listed with details.

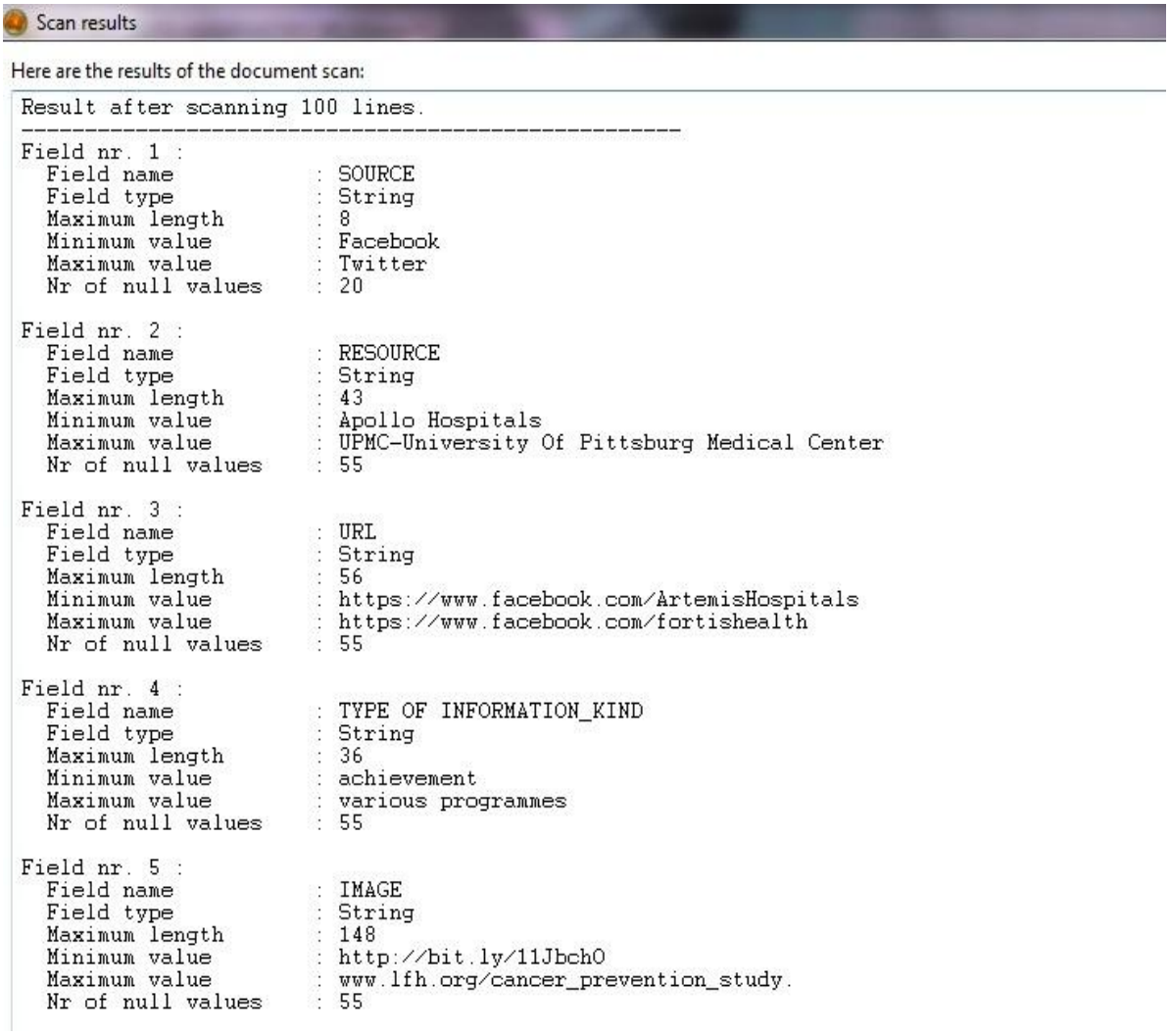


Figure 17 - Fields after analyzing the sample

One also has an option of viewing logging text that looks like:

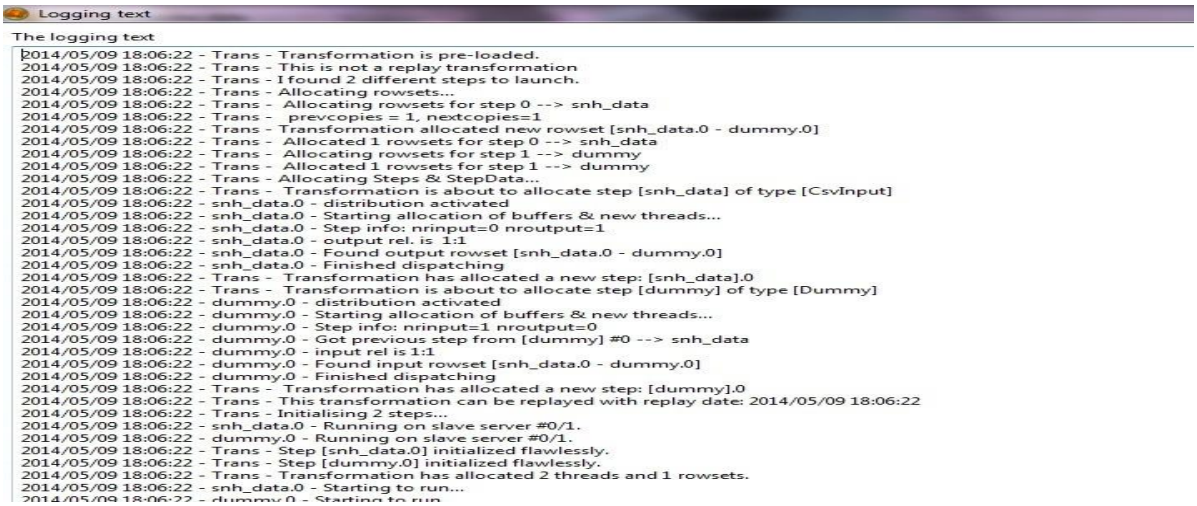


Figure 18 - Viewing logging text

Spoon also provides us with the option to preview our data. After this we need to set fields in the target file such as name of the data base and the table where the data will be loaded.

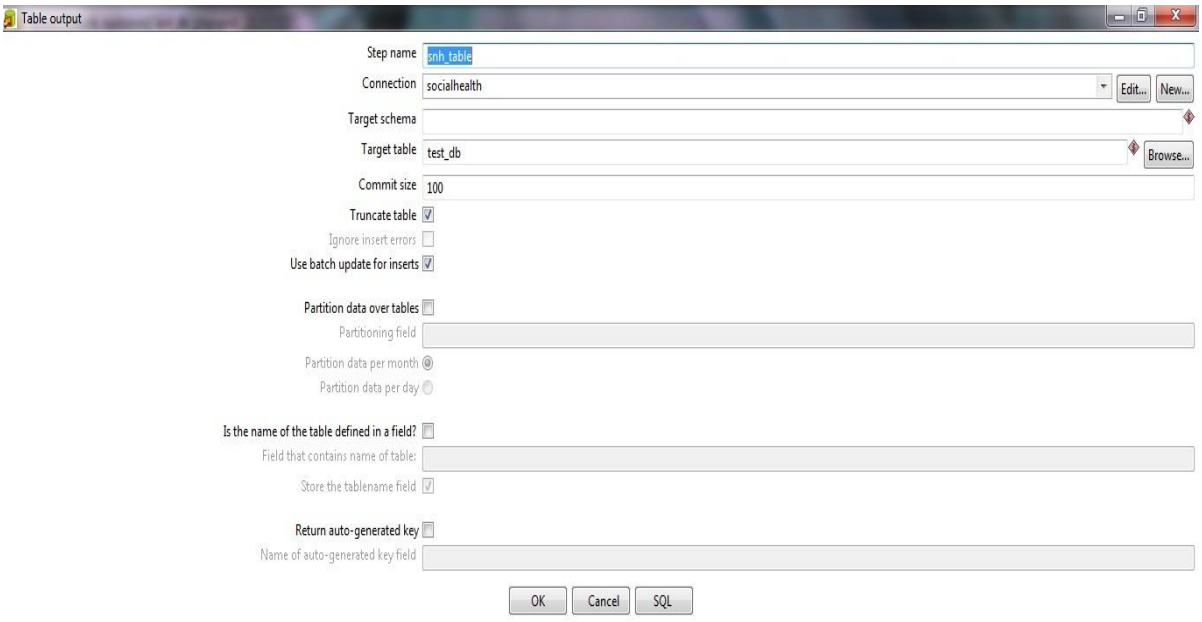


Figure 19 - The target description

Here the SQL button will result in altering my table snh_table depending on the attributes selected in the previous get fields.

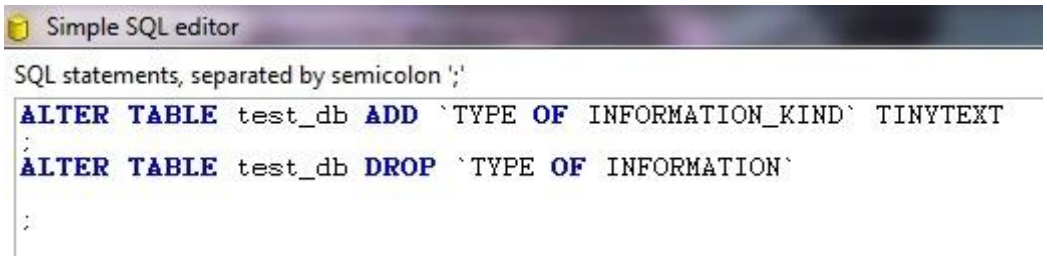


Figure 20 - Altering of the table in MySQL from Spoon

After the design has been completed we need to save it in the repository and then we can execute the transformation. In case if there is any error with the design it will throw an error, otherwise the data will be processed and records will be stored in the appropriate table in the database based on rules defined.

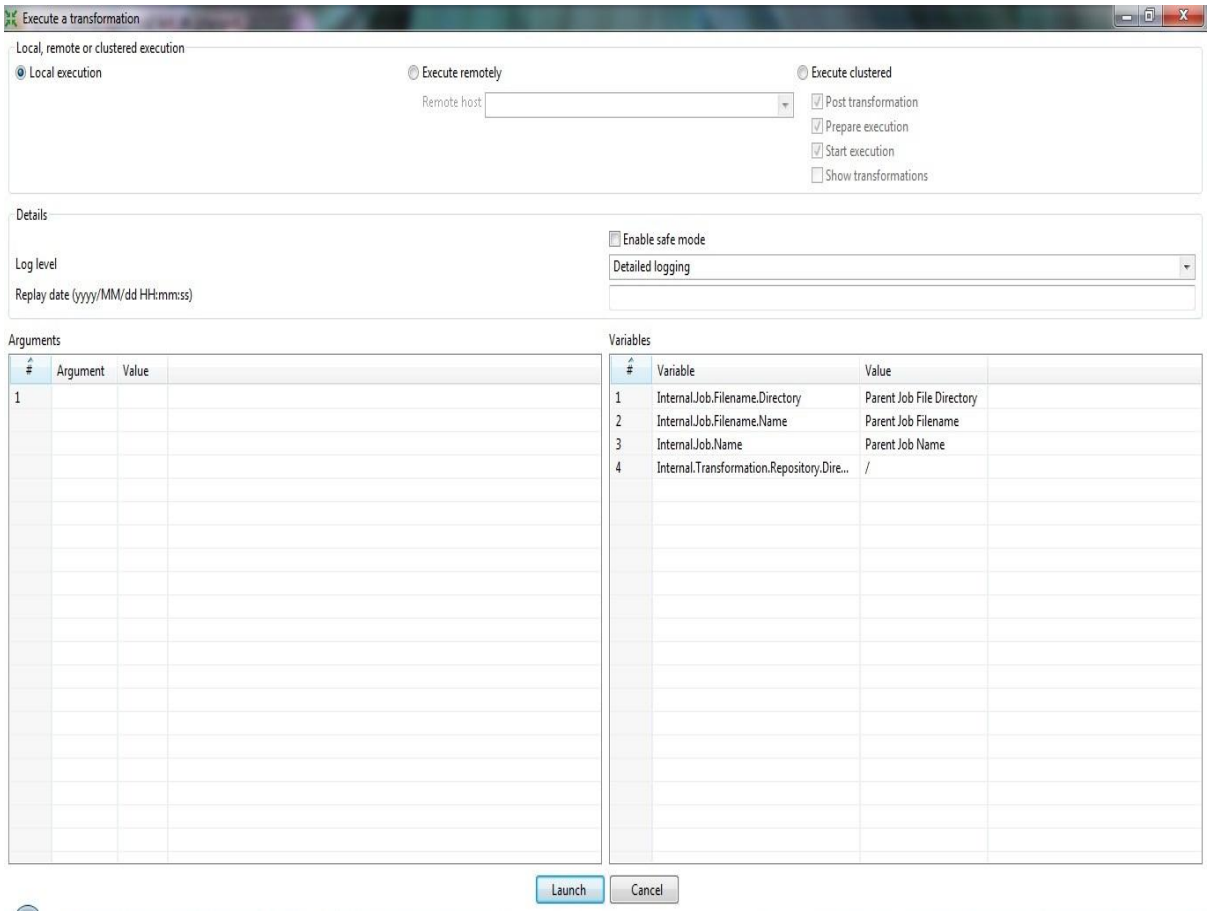


Figure 21 - Executing a transformation

After clicking the “launch” button, we get the execution results which is as follows:

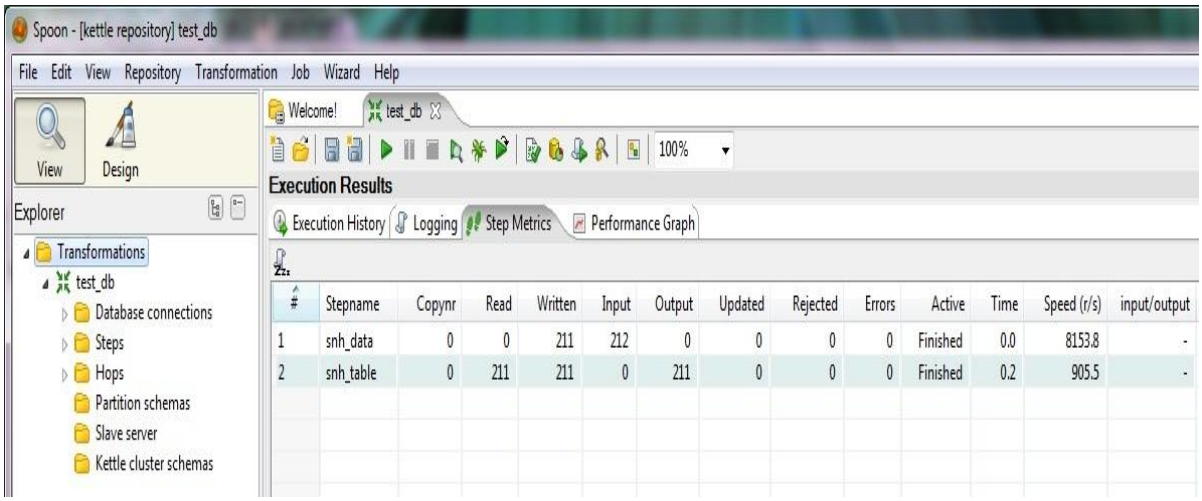


Figure-22 Execution results

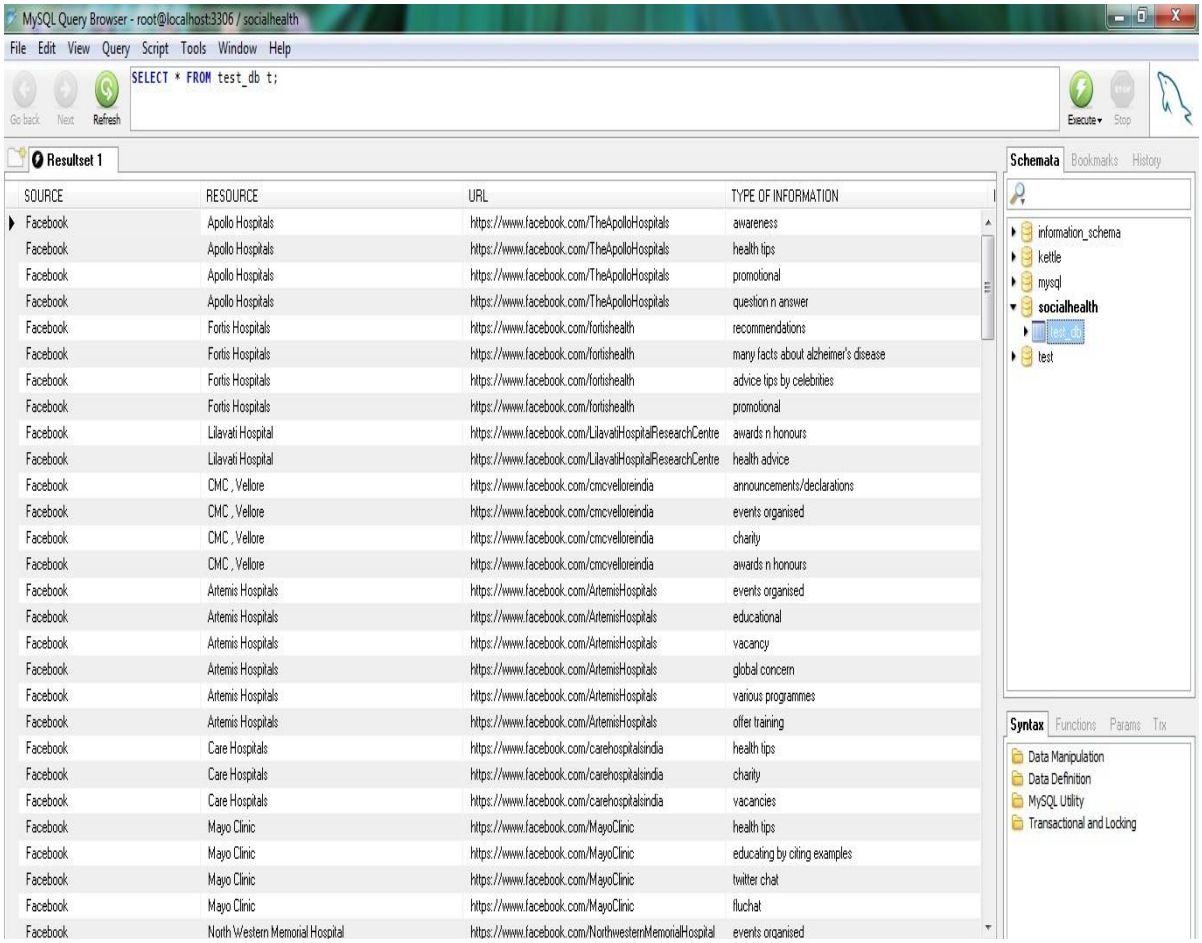


Figure-23 Table view in MySQL

3.2.4 Ontology Design

Once the database is created the next task that comes into picture is defining *ontology*. For this purpose I have written a code (program) which will calculate the frequency of the words appearing in the database and also it will calculate the transition probabilities of each and every word. The probabilities are arranged in descending order. The code is as follows:

```
#!c:/perl/bin/perl.exe
print "Content-type: text/html; charset=iso-8859-1\n\n";
use DBI;

#definition of variables
my (%count,$i,$key,@keys,$token,@tokens,@values);
my $db="socialhealth";
my $host="localhost";
my $user="root";
my $password="root";
my $total=0;

#connect to MySQL database
my $d   = DBI->connect ("DBI:ODBC:MySQL_SH")
                    or die "Can't connect to database: $DBI::errstr\n";

my $sth = $d->prepare("select * from test_db");
$sth->execute();
print "Number of rows found :" + $sth->rows;
print "\n";

while (my @row = $sth->fetchrow_array())
{
   sub tokenize
  {
  $_ = $_[0];
```

```
  s/\s+/\n/g;
  s/^\n//;
  s/$/\n/;
  s/([.,!?:;,])\n/\n$1\n/g;
  s/\n(["'`])([^\n])/\n$1\n$2/g;
  s/([^\n])(["'`])\n/$1\n$2\n/g;
  s/([^\n])([.,])\n/$1\n$2\n/g;
  s/\n([A-Z])\n\./\n$1./g;
  s/\n\.\n([^"A-Z])/\.\n$1/g;
  s/(\.[A-Z]+)\n\.\n/$1.\n/g;
  s/([^\n])'s\n/$1\n's\n/g;
  s/([^\n])n't\n/$1\nn't\n/g;
  s/([^\n])'re\n/$1\n're\n/g;
  s/\n\$([^\n])/\n\$\n$1/g;
  s/([^\n])%\n/$1\n%\n/g;
  s/Mr\n\.\n/Mr.\n/g;
  return(split(/\n/,$_));
 }

my $h=open(OUTFILE,">freq.txt") or die("cannot open output file");
 print OUTFILE"@row\n";
  $tokens = &tokenize(@row);
  foreach $token (@tokens)
{
        if ($token =~ /[a-zA-Z]/)
        {
                $count{$token} = $count{$token} ? $count{$token}+1 : 1;
        }
  }
}
$d->disconnect
or warn "Disconnection failed: $DBI::errstr\n";
```

```
# we need to sort a hash but we can only sort lists
# so put hash keys in list and sort the list
@keys = keys %count;
@keys = sort { $count{$b} <=> $count{$a} } @keys;
my $h=open(OUTFILE,">freq2.txt") or die("cannot open output file");
for ($i=0;$i<=$#keys;$i++)
{
        $total=$total+$count{$keys[$i]};  #calculation of transition probabilities
}
print "$total\n";
for ($i=0;$i<=$#keys;$i++)
{
        $values[$i]=$count{$keys[$i]}/$total;
        print OUTFILE "$count{$keys[$i]} $keys[$i]\n";
         print OUTFILE "$values[$i] $keys[$i]\n";
}
close(OUTFILE);
exit(0);
```

3.2.5 Crawling Algorithm

The algorithm proposed in the study will be crawling across the various social networking sites and retrieving prioritized results to the end user. The proposed approach is based on markovian approach. Steps followed in the markovian approach are as follows:

1. From the search phase, select the health/disease keyword and define it as variable Y.
2. Define the associate words in the phrase with the keyword 'Y' as X (x1, x2...xn).
3. Calculate the transition probability (T_p) of associated words with the keyword in the phrase based on the defined ontology.

$$\mathbf{T_p = P(X|Y) = P(X \cap Y) / P(Y)}$$

   Based on T_p observed, the results are prioritized in descending order and appropriate social networking information is retrieved.

3.2.6 Design of Webpage

To serve the above mentioned purposes I have designed a webpage for the same.

Figure-24 Snapshot of Web page (Index page)

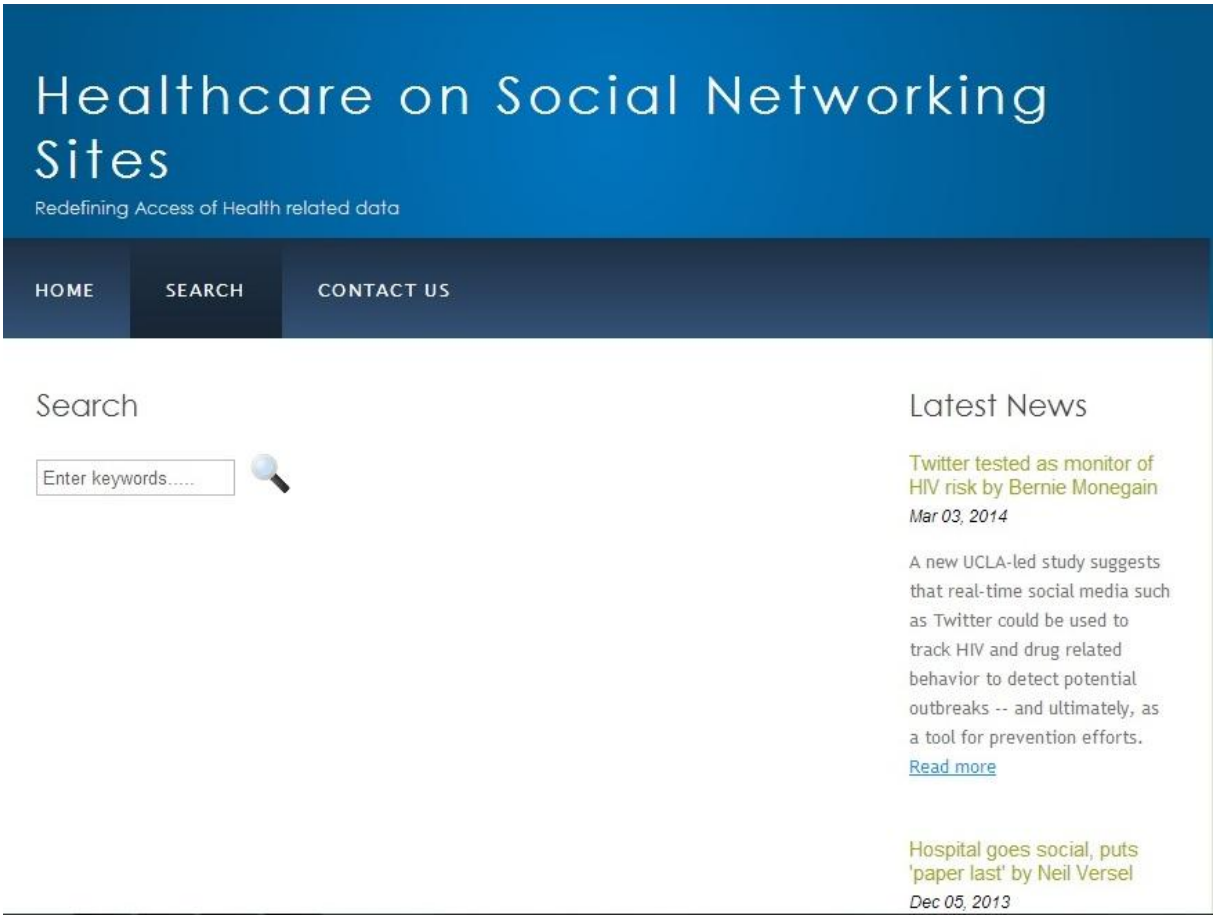


Figure-25 Snapshot of Search phase in the webpage

This is how our SEARCH option will look like in the webpage designed. As soon as the user types a keyword in the SEARCH phase, the terms associated with the given keyword will appear in descending order of their transition probabilities and accordingly information from appropriate social networking site is retrieved.

3.2.7 Deliverables

- Providing an integrated resource for accessing healthcare data available across different social sites - It would be beneficial for the end users as well as clinicians/healthcare organisations, as information will be available to them on a single go.
- Suggest the most optimal solution to end-user for their medical needs like doctor consultation, hospital visit, available treatment options, etc. corresponding to a particular geographical location.

Chapter - 4

CONCLUSION

Over the past few years, the healthcare landscape has evolved — to Health 2.0, wherein patients have assumed the role of active participants and seek control of their own health. Individuals are now interacting on social media websites and healthcare portals, seeking to reduce the communication gap between patients and healthcare providers. This rapid transformation is making healthcare delivery more patient-centric, and empowering users to make informed decisions about their health. Further, as doctors begin dealing with more informed patients, the time spent on patient care and the related costs will likely decrease, leading to improved overall standards of care. Better self-management of one's health could also help governments to control the growing healthcare costs and deal with the problem of staff shortages in the healthcare sector. Although some caution is necessary to deal with concomitant challenges posed by the use of social media in healthcare, a rational strategy adopted by healthcare organizations and regulatory agencies can help the industry to adequately respond to this important change in the healthcare landscape and fully leverage the opportunities offered by it.

References

[1] Dr. Vikram, "Impact of Social Media on Healthcare," Wipro Technologies, Bangalore, India, November 2010.

[2] Frances Griffiths a, Jonathan Cave a, Felicity Boardman a, Justin Ren b, Teresa Pawlikowska a, Robin Ball, Aileen Clarke a, Alan Cohen, "Social networks - The future for health care delivery," University of Warwick, Coventry, West Midlands CV4 7AL, UK, September 1, 2012.

[3] Cartledge P, Miller M, Phillips B, "The use of social-networking sites in medical education," Med Teach. 2013 Oct; 35(10):847-57.

[4] Dietrich F, Shipherd AM, Gershgoren L, et al., "Sport psychology group consultation using social networking web sites," Psychol Serv. 2012 Aug;9(3):323-4.

[5] Omurtag K, Jimenez PT, Ratts V, et al., " The ART of social networking: how SART member clinics are connecting with patients online," Fertil Steril. 2012 Jan; 97(1):88-94.

[6] Jaganath D, Gill HK, Cohen AC, Young SD, "Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention," AIDS Care, 2012; 24(5):593-600.

[7] George DR, "Making 'social' safer: are Facebook and other online networks becoming less hazardous for health professionals?," J Clin Ethics. 2012 Winter; 23(4):348-52.

[8] Phil Baumann, 140 Health Care Uses For Twitter, Internet : http://philbaumann.com/140-health-care-uses-for-twitter [Jan 16,2009].

[9] Twitter doctors directory. [Internet : http://twitterdoctors.net/]

[10] Should I Be "Friends" with My Patients on Social Networking Web Sites? [Internet: http://www.aafp.org/afp/2011/0701/p105.html]

[11] Deb B, Srirama SN, "Social networks for eHealth solutions on cloud," Front Genet, 2013 Sep 3;4:171.

[12] Pentaho Corporation. Pentaho Data Integration (Kettle). Internet: www.kettle.pentaho.com,[Jan 15,2011].

[13] UniSIM HTML, Internet:
sst.unisim.edu.sg:8080/dspace/bitstream/123456789/171/1/09_Fan%20Ying%20Xin.pdf,[Jan 15,2011].

[14] Perl and CGI Tutorial. [Internet: http:// http://www.tutorialspoint.com/perl/perl_cgi.htm]

[15] Rishi Kanth Saripalle, "Current status of ontologies in Biomedical and Clinical Informatics," University of Connecticut, Stors , Journal of Biomedical Informatics (37) 2004.