Towards Personalized Health Profiling in Social Network

Rabia Batool*, Wajahat Ali Khan, Maqbool Hussain, Jahanzeb Maqbool, Muhammad Afzal and Sungyoung Lee
*Department of Biomedical Engineering**
Department of Computer Engineering
Kyung Hee University
Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do, Republic of Korea, 446-701
{*rabia, wajahat.alikhan, maqbool.hussain, jahanzeb.maqbool, muhammad.afzal, sylee*}*@oslab.khu.ac.kr*

***Abstract*—E-Health industry is expanding considerably due to its effectiveness and efficiency in provision of better healthcare to people. The concern to get personalized information at one's doorstep is on rise and social media is filling this gap by connecting users and sharing information. Personalized healthcare information in social media has also become the cornerstone allowing us to investigate innovative ways to effectively utilize its needs. The proposed approach uses personalized healthcare concepts by analyzing user posts on social media network. The analysis indicates the discrepancies in managing ones health and provide guidelines to overcome them. Methodology includes building user profile by identifying user related topics, emotions and health condition. Our experiment investigated 6000 tweets for diabetes. After filtering we managed to run our experiment on 4000 health related tweets. Smart Clinical Decision Support System (smart CDSS) [1] service utilizes the results generated by our experiment.**

***Keywords*-User Modeling; Personalization; smart CDSS; Tweets; Social Media; Healthcare;**

I. Introduction

World has turned into a global village due to the drastic increase of information sharing through media. The rise of social media like Facebook, MySpace and Twitter in less than a decade has changed the general perspective of networking, socialization and personalization. These social networks have significant impact on the daily life of millions of users. Users having common interests, activities and language, interacts and share their racial, sexual, religious and demographic identities [2].

Shifting of real time information in healthcare era helps in providing more immediate interventions in the care provided to individuals in the hospital, at clinic or at home. Social media contribution towards healthcare is playing a significant role in real time healthcare provisioning. For example, research [3] investigated that four out of five users are using internet to find out personalized healthcare information related to the particular disease and its treatments. Individuals can use social media to examine other's experience in terms of symptoms, reactions and treatments about particular disease. Healthcare organizations can also take benefit by finding timely response of problems and monitoring user's behaviours, conditions and feelings in-between the visits. Social media empowers users to know more about themselves including their health. To disseminate healthcare related awareness, people will be more prepared to manage minefield of modern medical treatment. The problem lies how patients can effectively use social media to manage their health related issues. Also how personalization can be ensured by exposing personal health related issues.

With internet and technology, communication has become simple and fast. Due to these simplicity trends, healthcare is also moving towards information sharing and communication. Paul et al. study online social networks to analyze their effect on patient health and found that people get more benefit while sharing their data on social networks like patientlikeme portal [4]. These benefits can be medication regularity, reduction in treatment side effects, knowing others having same symptoms and better healthcare from professionals. Therefore people who expose their data with more friends get more benefits [5]. According to research report by PwC Health Research Institute [6] almost 80% users in the age of 18-24 were willing to share their health information on social networks. Social networks will peel back every corner of the health system and drive transparency on cost, value, and outcomes [7]. Personal post on social media can help to complete patient profile. 18% of user who use social media for healthcare track and share health symptoms and 24% are those who share their health experiences with others [7]. People are using information from social media to make decision in healthcare domain. According to report 45% of consumers accept that social media would effect their decision to go for hospital visit. More than 40% of users said that social media helps in the way they coped with a chronic condition, their approach to diet and exercise, and even in searching for doctor [7].

The proposed system monitors changing health conditions, emotions and interests of users and provides real time personalized health information and guidance from highly qualified physician's knowledge. User interests and emotions are extracted by investigating user's social data which is passed to smart CDSS for decision making. The response is in the form of approved recommendations or alerts, based on stored clinician's knowledge, to the user/patient. Decision from smart CDSS is personalized information based on user's interest and health condition. This system improves

healthcare activities, self-management of care and enhances communication for patients and physicians. The rest of paper is structured as follows.

Section II discusses the related work closely aligned with our work. Section III highlights the importance and relationship of healthcare social data and decision making. Section IV describes the proposed system architecture and its components in detail. Section V explains the proposed system working based on a real scenario. In Section VI, discussion about the necessity of using social media for healthcare purposes is highlighted. At the end, section VII concludes the work and provides future directions.

II. Related Work

Social media is attracting the attention of researchers due to the availability of useful information openly. In [8], research study is carried out about information diffusion pattern on Twitter. It analyzed structure of Twitter network to identify influential users for viral marketing. J. Chen et al. analyzed URL recommendations on Twitter using data stream technique [9]. Methodology of the system is based on content sources, topic interest models and social voting to design URL recommender and compare different recommender techniques. Fabian Abel et al. analyze user modeling on Twitter for personalized news recommendation and enrich news with tweets to improve the semantic of Twitter activities [10] [11] [12]. The work is based on the analysis of user modeling techniques including topic based, entity based and hostages based. Also focus is on temporal pattern extraction in user profile. Ilknur Celik et al. studied semantic relationship between entities in Twitter to provide a medium where users can easily access relevant content for what they are interested in [13]. The research work consider certain time period to analyze performance in learning relationship. It shows that for those types of relationships Twitter is a suitable source as it allows for discovering trending topics with higher accuracy and with lower delay in time than traditional news media [13]. Lu et al. proposed a new ranking model for personalized local search [14]. The model is based on personalized search functionality by integrating Twitter social network structure and content analysis. Laniado and Mika studied the use of hashtags on Twitter and showed that about fifty percent of the hashtags can be mapped to entities in freebase [15] .The system is categorized in to four dimensions: frequency, specificity, consistency and stability to assess hashtags as strong identifiers [15]. Matchlock and David Muller proposed a method to associate hashtags with encyclopedia entities [16]. System uses Wikipedia entities as description of hashtags in microblogging service to understand the actual context of hashtag.

There exists a problem of insufficient user data for efficient recommendation. Balabanovic et al. proposed a system to build user profile by combining both collaborative and content based recommendation techniques [17] . In content-based recommendation systems, user preferences are considered for providing recommendations. On the other hand, in collaborative recommendation the system identifies users whose tastes are similar to those of the given user and recommends items they have liked [17].

In this paper we go beyond profiling of users by also providing the users with personalized health related recommendations and alerts, based on stored clinicians knowledge. We propose a novel approach for analyzing the behavior and lifestyle of individuals by monitoring patient self reported data and social posts to provide recommendations for better healthcare.

III. HEALTHCARE, TWEETS AND DECISION MAKING

By monitoring person's social activities, interest and emotions can be extracted, helping clinicians to provide better guidance to patients. System provides alerts when alarming health condition occurs, allowing patients and physicians to intervene and modify treatment plans as needed. Our proposed system integrates as a plug-in application, while extracting user related information including profile information, person interests and emotions. Analyzing the interest, behaviour and lifestyle of person provides assistance in better decision making. For example, depression and violence symptoms can easily be identified from a person's social interests and emotions. Alchemy API [18] is used for the extraction of user's interests from free text. This API accepts unstructured text, processes it using natural language processing and machine learning algorithms, and returns entities, sentiment of users about specific entity and also concepts related to those entities. *Adaptor*, a cloud based application works as a connector, performs automatic mapping of this knowledge into virtual Medical Record (vMR) [19]. A vMR for Clinical Decision Support (CDS) is a data model for representing individual medical records and clinical information inputs and outputs that can be exchanged between CDS engines. In order to

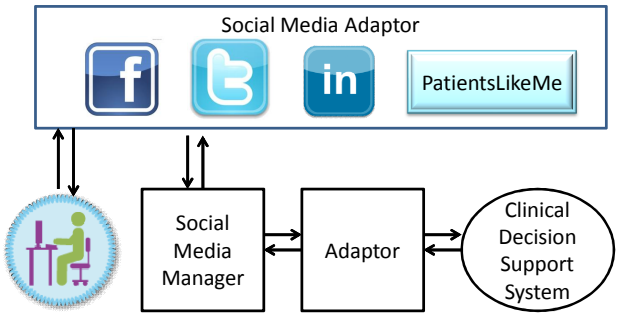

Figure 1. Work flow of Smart H-Tweet Engine

achieve interoperability, SNOMED CT which is the most comprehensive, multilingual clinical healthcare terminology in the world [20] is used as vocabulary to coded attributes. vMR is further processed by smart CDSS to provide better recommendations to a person/patient. Physicians verifies the recommendations provided by smart CDSS. This data can also be used by clinicians to know more about patients activities. Fig 1 shows the work flow of whole system.

IV. PROPOSED SYSTEM ARCHITECTURE

The proposed system uses two distinct technologies (1) Natural Language Processing on user's post to extract knowledge and (2) HL7's vMR for representing the clinical term. Architecture consists of two main components as shown in Fig 2;*Social Media Manager* and *Adaptor*. These components are elaborated as follows.

A. Social Media Manager

Social Media Manager is plugable application that interacts with social media. It requires a lot of processing on free text from social media. *Social Media Manager* consists of the following subcomponents.

1) Data Manager: This component is responsible for fetching data from social media and processing the fetched data. It has two parts.

- Data Fetcher:
  Data Fetcher sends request to social media for stream of user. The fetched data is in different format for each media.
- Data Processor:
  Fetched data requires some pre-processing before analyzing. Users use abbreviations to save time and space such as Twitter user can post up to 140 character in one tweet. Such kind of noise in data effects knowledge extracted from tweets. Therefore to remove such kind of noise, *Data Preprocessor* first removes slang and abbreviated word using slang lexicon. We have lexicon of 1300 slang commonly used in social media posts and chats. Twitter data also consist of lots of spelling mistake. Lots of works has been done on spelling correction so spelling correction is not in the scop of this paper. *Data Processor* then filters posts which can be useful in knowing user health conditions and converts filtered post in required useable format. If post contains a URL, it separates URL from text to process text and URL separately.

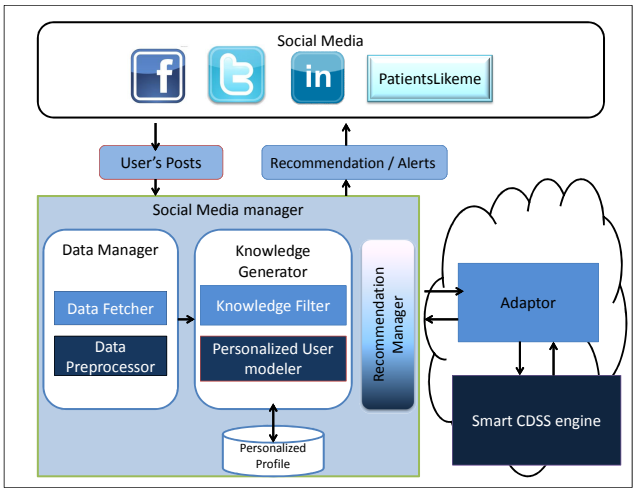

Figure 2. Architecture of Smart H-Tweet Engine

2) Knowledge Generator: Data Manager Component forwards the pre-processed data to *Knowledge Generator*. The main purpose of this component is to extract user's interests, emotions and health conditions from social network data and maintain history to build personalized profile. Following is the detail of *Knowledge Generator's* component.

- Knowledge Filter:
  Knowledge Filter extracts user's interest by using Alchemy API. It obtains knowledge by exposing the semantic richness hidden in post using named entity extraction, keyword extraction, and concept tagging. *Knowledge Filter* also extracts user's health condition if user specifies it in posts and user's emotions based on keywords used in user's post. To extract emotional keywords from posts, system uses emotional keywords dictionary which is used to identify emotion related to different keywords. API also provides sentiment related to each entity which provide context based information. System also performs part of speech tagging to identify structure, semantics and relations of extracted knowledge with users. After processing by *knowledge Filter*, data is converted into meaningful information and usable knowledge.
- Personalized User Modeler: Better decision making requires maintenance of history of individuals interest and conditions. *Personalized User Modeler* maintains user's data in Personalized Profile and pass this information to smart CDSS and clinicians when required.
- Personalized Profile: This database maintains users social network interest and their health condition history to provide them better guidance and information at right time.

B. Adaptor

Adaptor is cloud based application used to send structured data to smart CDSS. The unstructured data received from *Social Media Manager* is transformed into format that confirms standard interface of smart CDSS. Data *Adaptor* converts user's personal profile to vMR format and sends it to smart CDSS. After processing from smart CDSS, the recommendations and alerts are generated and finally converted to user readable format by *Adaptor*.

Number	Tweet	Keyword	Category	Concept	Entity		
					Type	Text	Sentiment
1	I feel my high blood pressure is at an unsafe level every time I'm at work. It's seriously going to give me a depression one of these days.	High blood pressure	Health	Hypertension, Orthostatic, hypotension, Blood pressure	Health condition	Depression	Negative
					Health condition	Blood pressure	Negative
2	I am Diabetic. Here's how it works. My insulin pump and continuous glucose meter (CGM). Plzzz help me	Glucose meter, Insulin pump, Diabetic	Health	Insulin, Diabetes mellitus, Glucose, Hypoglycemia, Diabetes, Glucose meter,	Health condition	Diabetes	Negative
3	Wide awake, I've got a headache & work in the morning. .	Wide awake	Health		Health condition	Headaches	Negative
4	I am healthy and feeling good after having high blood pressure now	Blood pressure	Health	Hypertension, Orthostatic, Hypotension, Blood pressure,	Health condition	High blood pressure	Positive
5	I thought I was in dream and in reality I was in a coma.	Dream	Health	Mind Psychology	Health condition	Coma	Negative
6	I really hope my glucose test goes ok. If I have gestational diabetes II .I be sad. I love my sweets lately.	Gestational diabetes , Glucose test, Sweets	Health	Diabetes mellitus, Obstetrics	Health condition	Diabetes	Negative

Figure 3. Knowledge extracted from tweets

```
<vmrInput>
  <templateId root="2.16.840.1.113883.3.795.11.1.1"/>
  <patient>
    ……
    <clinicalStatements><!--Health problem>
      <problems>
        <problem> <!-- High blood pressure-->
          <id root="107c2dc0-67a5-11db-bd13-0800200c9a66"/>
          <problemCode codeSystem="2.16.840.1.113883.6.96"
            codeSystemName="SNOMED CT" code = "170581003" />
        </problem>
        </problems>
      <Observations>
        <Observation>
          <substance> <!—Depression-->
            <id root="2d813912-a3d1-464c-9567-4d92c75a0c23"/>
            <substanceCodecodeSystem="2.16.840.1.113883.6.96“
            codeSystemName = "SNOMEDCT"code ="310495003 “ />
          </substance>
        </Observation>
      </Observations>
    </clinicalStatement>
  </patient>
</vmrInput>
```

Figure 4. Standardized input for smart CDSS

1) Standardized input for smart CDSS: SNOMED CT is used to convert personal interest and health condition into semantic codes that are recognized by smart CDSS. Knowledge extracted from user post becomes input for *Adaptor*. *Adaptor* performs intelligent mapping of personal data to the coded standard vMR format as there are many codes for one disease under different conditions. Fig 4 shows example of vMR input build using user's personal interest from tweet number 1 as shown in Fig 3. Data from tweet 1 can be represented as problem and observation clinical statement.

V. A CASE STUDY

This section provides proof of concept by elaborating case study reflecting the working environment of whole system. Our initial target implementation of the whole system addresses recommendations and alerts for diabetic patient. For case study we consider Twitter data as input and test our system for diabetic patients. We analyzed Twitter public data and collected around 6000 tweets related to diabetes. Some of the tweets collected is shown in Fig 5. It shows tweets related to information of user's blood pressure. *Data Fetcher* sends request to Twitter to collect tweets. We suppose that tweets are from the same user whose clinical data is available with smart CDSS. Data Preprocessor format data and observes URL use in the tweet. Due to free text

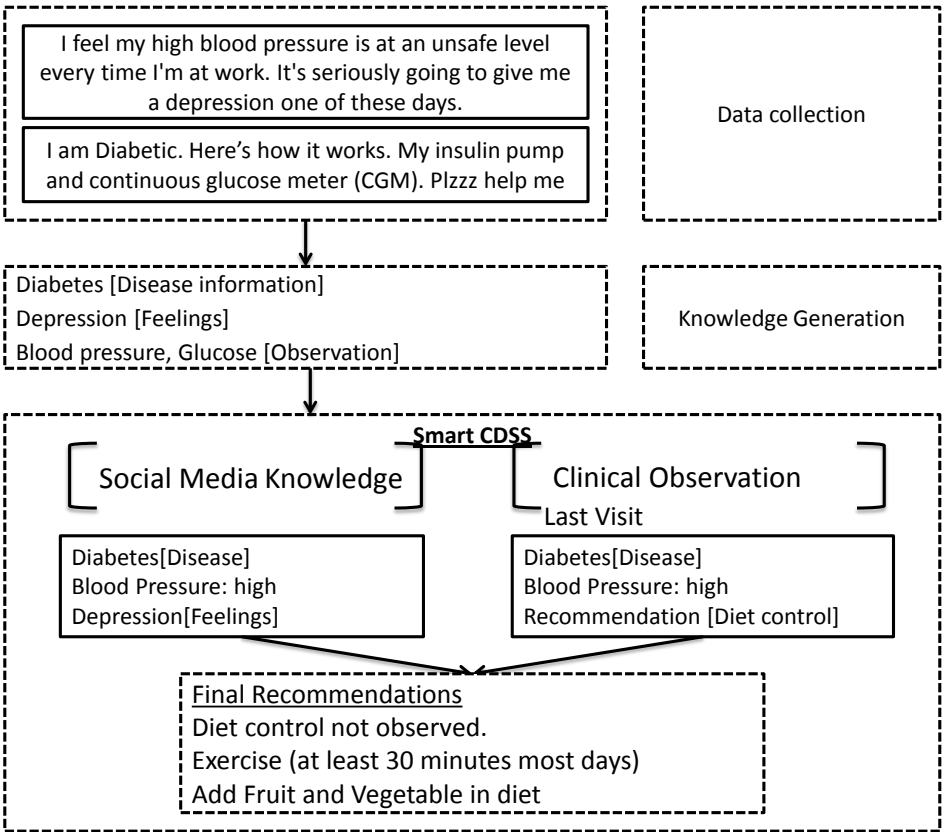

Figure 5. Tweets, knowledge and resulting recommendations

Twitter users use slang or abbreviations. *Data preprocessor* component processes tweets (6000 tweets in our scenario) to identify presence of any short hand notation and normalizes them to meaningful words. It also checks for any repetition of same character more than two times. Some of these includes 'plz' for 'please' and 'goood' for 'good'. *Data preprocessor* populates tweet object for each tweet and forwards data to *Knowledge generator*. *Knowledge filter* extracts knowledge from tweets and collects user's emotions, health conditions, interests and category of tweets. Using Alchemy API, Knowledge Filter collect keywords, concepts, entities, and sentiments of user for a particular entity in tweets. The table in Fig 3 shows knowledge extracted from some tweets. With the help of Alchemy API, we extracted keywords, concepts, entity and tweet category from user's tweets. Moreover, table shows type of entity, subtypes to overcome ambiguity and the sentiments of user about that entity. Before loading data into database *Personalized User Modeler* component classify tweets and filter all tweets lying in health category and discard all other tweets. After filtering system gets around 4000 tweets from health category. *Adaptor* collects filtered and processed information and provide it to smart CDSS in its compliant standardized vMR format. Smart CDSS processes tweet based user's health information in accordance to his/her medical information. Recommendations are generated by comparing and analyzing this data with clinical observations in the medical record. Final recommendations or guidelines are provided to the clinicians or users who can take advantages of the information and modify health plans accordingly. As smart CDSS pass the recommendations and alerts verified from clinicians, so there is no fear to get wrong recommendations. Fig 5 explains the processed data and resulting recommendations for single user.

VI. DISCUSSION

Increase in the use of social media also increase the importance for issues directly concerning users like data privacy and security. Systems are available that are using data residing on social media but the main issues of security and privacy needs to be tackled. The integration of healthcare with any other system increases the need of privacy and security by multiple times. Annonymization technique is one solution to handle security and privacy concerns. Patient demographics related data should be annonymized to secure this approach. Accuracy and validation of information is another concern for such systems. One cannot completely rely on the correctness of information provided by users.

Therefore clinicians validation of information is necessary before reaching any decision.

For effective healthcare provision, integrating smart CDSS with social networks will enable users receiving more information and guidelines increasing their level of awareness. This will also result in reduced healthcare costs for hospital resource consumption and provide timely personalized healthcare services to patients. Organizations and physicians will be able to record and monitor their patient in-between the visits through social media. Proposed system provides right care to the right person at right time by keeping user's data confidential and secure.

VII. CONCLUSION AND FUTURE WORK

In this paper we demonstrated a system to extract user interests, health conditions and emotions from social media for creating semantically rich user profile. This provides the users with personalized information and guidelines and more knowledge to clinicians about patients. The proposed system opens new dimensions to effectively utilize data generated using social media for healthcare. It focuses on efficient provision of health related services by utilizing social data effectively. The work also highlights some key challenges like provisioning of health services by taking account of security and privacy concerns. We also have plan to find relationship between different user's posts to detect effectiveness of guidelines and alerts. By using some more techniques on user posts to retrieve some additional textual information will also help in knowledge discovery for personalized recommendations and alerts.

ACKNOWLEDGEMENT

This research was supported by the MKE(The Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2012-(H0301-12-2001)).

This work was supported a grant from the NIPA(National IT Industry Promotion Agency) in 2012. (Global IT Talents Program).

REFERENCES

[1] M. Hussain, W. A. Khan, M. Afzal, and S. Lee, "Smart cdss for smart homes," in *Proceedings of 10th International Conference On Smart homes and health Telematics*, 2012.

[2] danah m. boyd and N. B. Ellison, "Social network sites: Definition, history, and scholarship," (Last visited in March 2012). [Online]. Available: jcmc.indiana.edu/vol13/issue1/boyd.ellison.html

[3] S. Fox, "Health topics," (Last visited in March 2012). [Online]. Available: pewinternet.org/Reports/2011/HealthTopics.aspx

[4] "Patientslikeme," (Last visited in August 2012). [Online]. Available: www.patientslikeme.com

[5] P. Wicks, D. L. Keininger, M. P. Massagli, C. de la Loge, C. Brownstein, J. Isojarvi, and J. Heywood, "Perceived benets of sharing health data between people with epilepsy on an online platform," *Epilepsy and Behavior*, pp. 16–23, 2012.

[6] "Health research institute," (Last visited in august 2012). [Online]. Available: http://www.pwc.com/us/en/health-industries/health-research-institute/index.jhtml

[7] H. R. Institute, "Social media likes healthcare, from marketing to social business," (Last visited in august 2012). [Online]. Available: http://www.pwc.com/us/en/health-industries/publications/health-care-social-media.jhtml

[8] M. Cha, H. Haddadi, F. Benevenuto, and K. Gummadi, "Measuring user influence in twitter:the million follower fallacy," in *proceeding of International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2010.

[9] J. Chen, R. Nairn, L. Nelson, M. Bernstein, and E. Chi, "Short and tweet: experiments on recommending content from information streams," in *Proceedings of the 28th international conference on Human factors in computing systems*. ACM, 2010, pp. 1185–1194.

[10] F. Abel, Q. Gao, G. Houben, and K. Tao, "Analyzing user modeling on twitter for personalized news recommendations," *User Modeling, Adaption and Personalization*, pp. 1–12, 2011.

[11] F. Abel, Q. Gao, G.-J. Houben, and K. Tao., "Semantic enrichment of twitter posts for user profile construction on the social web," *The Semanic Web: Research and Applications*, pp. 375–389, 2011.

[12] F. Abel, Q. Gao, G. Houben, and K. Tao, "Analyzing temporal dynamics in twitter profiles for personalized recommendations in the social web," in *Proceedings of ACM WebSci '11, 3rd International Conference on Web Science*. ACM, 2011.

[13] I. Celik, F. Abel, and G. Houben, "Learning semantic relationships between entities in twitter," *Web Engineering*, pp. 167–181, 2011.

[14] L. Guo and M. Lease, "Personalizing local search with twitter," in *proceeding of SIGIR Workshop on Enriching Information Retrieval*, 2011.

[15] D. Laniado and P. Mika, "Making sense of twitter," *The Semantic Web–ISWC 2010*, pp. 470–485, 2010.

[16] U. Lösch and D. Müller, "Mapping microblog posts to encyclopedia articles."

[17] M. Balabanović and Y. Shoham, "Fab: content-based, collaborative recommendation," *Communications of the ACM*, vol. 40, no. 3, pp. 66–72, 1997.

[18] "Alchemy api," (Last visited in March 2012). [Online]. Available: www.alchemyapi.com

[19] "Implementation of virtual medical record object model for a standards-based clinical decision support rule engine," (Last visited in March 2012). [Online]. Available: www.ncbi.nlm.nih.gov/pmc/articles/PMC1839394/

[20] "Snomed," (Last visited in March 2012). [Online]. Available: www.ihtsdo.org/snomed-ct