Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences

Procedia - Social and Behavioral Sciences 195 (2015) 1990 – 1998

World Conference on Technology, Innovation and Entrepreneurship

State of the Art of Semantic Web for Healthcare

Xhemal Zenuni[a]*, Bujar Raufi[a], Florije Ismaili[a], Jaumin Ajdari[a]

[a]*Faculty of Contemporary Sciences and Technologies, SEE University, bul. Ilindenska nn, 1200 Tetovo, R. of Macedonia*

Abstract

The ultimate goal to improve healthcare practices and the development of better biomedical products largely depends on the ability to share and link the wealth of collected medical data. The key challenge to pursue this ambitious objective is not only enabling the integration of the data spanning heterogeneous data sources and formats, but in the development of tools and standards for flexible search, data analytics and user friendly interfaces. In this paper we conduct an extensive survey on how Semantic Web is used to answer these challenges. First, we review ontology management and semantic data repositories for healthcare. Second, we conduct a survey on most representative applications and user – friendly viewers for semantic healthcare data. Third, we analyze the data mining and data analytics approaches used to find useful patterns and knowledge in these data. Finally, we discuss the positive effects of this synergy between Semantic Web and healthcare processes, and we identify some of the major remaining obstacles and research challenges in this area.

Peer-review under responsibility of Istanbul Univeristy.

Keywords: Semantic Web, Healthcare, Semanctic Tools, State Of Art

1. Introduction

The computerization of health, such as the utilization of information systems and technological medical devices on daily basis in hospitals and other medical institutions have already produced and will continue to generate tremendous amount of data from medical records, patient monitoring and medical imaging. This explosive growth of health related data needs properly to be explored in order to uncover valuable information and to transform such data into valuable knowledge that could lead to improved healthcare practices and to the development of better biomedical products. Yet, this ultimate goal continues to be a daunting task due to many challenges.

*Corresponding author, Tel.: +389 44 356 185
E-mail: xh.zenuni@seeu.edu.mk

Peer-review under responsibility of Istanbul Univeristy.
doi:10.1016/j.sbspro.2015.06.213

One of the major challenges is interoperability of health and medical data. The generated data not only comes from different sources, but it is stored in different and distributed data repositories, it resides within different administrative domains and it has inconsistency in naming, structure and format. An important requirement is to capture relevant data, but also to make it widely available for others around the globe in a format that will be easily accessible, accurate and manageable in terms of efficient data processing and integration with other systems.

In addition to the data integration problem, the user interaction with the data is another challenge. The difficulty lies in search handling, data navigation and data presentation. Developing tools that will enable efficient finding of resources or data are difficult to achieve, especially when can be a sea of possible links and interconnections between the published data from many different sources. And such associations could be utilized in the background to provide new contexts and possibilities for more efficient search results. In addition, another important requirement is to provide consistent presentation of that data and to enable easy navigation and efficient data exploration, even from different devices.

Finally, another challenge is how to use this huge amount of data to find valuable new patterns and transform such data in to a valuable knowledge leading to potential improvement of resource utilization, patient health and the development of biomedical product. For example, a data mining application could explore patient data, such as symptoms, conditions or family history to detect the causes of diseases and/or to suggest medical solution to the patient at lower cost. There is a vast potential for data mining and data analytics tools in healthcare that could lead to useful information for decision making and generating such tools that are able to analyze and extract information from vast and complex health data is critical and complex requirement.

In recent years, Semantic Web is gaining ground to address these challenges. The World Wide Web Consortium has established the Semantic Web for Health Care and Life Sciences Interest Group whose mission is to develop, advocate and support the use of Semantic Web technologies in healthcare and related fields (HCLSIG, 2015). The aim of this paper is to analyze the current status on how Semantic Web is used to solve the health data integration and interoperability problem, how it provides advanced data linking capabilities that can improve search and retrieval of medical data, we overview the developed interfaces that allow efficient exploration of such data, and finally we analyze the tools and approaches to semantic health data analysis and knowledge discovery.

The remaining of this paper is organized as follows: Section 2 surveys the representative ontologies and the major available semantic data repositories developed in healthcare domain. Section 3overviews human – computer interaction interfaces and the search approaches to healthcare semantic data. Section 4 explores how semantic information is used in different data mining applications for the healthcare domain. Finally, in Section five we discuss the positive effects of this synergy between Semantic Web and healthcare processes, and we identify some of the major remaining obstacles and research challenges in this area. Section 6 concludes the work and we reflect on the likely impact and the future of this paradigm.

2. Ontologies and Semantic Data Repositories for Healthcare

Ontologies serve as a central component in the Semantic Web knowledge infrastructure. They provide controlled vocabularies of scientific terminology used to assist in annotation of produced data, such as the basic terms and relations in a domain of interest, and as well as rules how to combine these terms and relations. They are used to provide a common terminology over a domain and the basis for interoperability of systems. Consequently, we first survey the ontologies developed in healthcare domain. However, in addition to ontology, Semantic Web has developed standards and technologies, such as RDF to publish and link the data. Therefore, in the second part of this section we overview the available public health related datasets published using Semantic Web standards and technologies.

2.1. Ontologies

In the past years, several healthcare domain ontologies have been created. Most of them have been created to describe a specific domain in biomedicine, such as the terms to describe anatomical parts and their relations, or terms used in clinical medicine, such as in Electronic Health Records or rehabilitation domain. The list of ontologies is continuously growing and most of them are available via BioPortal (BioPortal, 2015) and some of the features of referential health and medical ontologies are summarized in Table 1:

- SNOMED CT has been developed over the years and is considered as the main ontology for representation of clinical concepts, terms and relationships in the field of health care (SNOMEDCT, 2015). The coding and classification work have been driven by three fundamental design criteria, namely the understandability, reproducibility and usefulness [1]. The ontology has hierarchy structure with a set of top level general concepts. All other concepts are subtypes of one these top concepts, and further down in hierarchy we move, the more specialized the concepts become. SNOMED-CT covers most of the areas that are used in medical practice, including clinical findings, symptoms, diagnoses, pharmaceuticals, body structures, medical devices, social contexts, and so on. Each concept is assigned a unique ConceptID and a Fully Specified Name(FSD),which can be interpreted as being a unique human – readable description of that concept. SNOMED-CT provides a consistent way for indexing, storing, retrieving and aggregating clinical data that can enhance the interoperability between different health information systems.
- Logical Observation Identifiers Names and Codes (LOINCS) provide a universal code system for laboratory test and other clinical observations (LOINC, 2015). For each observation, such as arterial blood gases, homogram or serum potassium, the database includes a code, a short name, a long formal name and synonyms. The aim of LOINCS is to provide common codes and terminology, so that when hospitals, pharmaceutical manufacturers, researchers, and public health departments although receive messages from multiple sources, they can easily or automatically file in the right slots of their medical records, research, and/or public health systems.
- MedDRA(Medical Dictionary for Regulatory Activities) is a rich and highly specific and standardized terminology to facilitate sharing of regulatory information internationally for medical products used by humans (MedDRA, 2015). Products covered by the scope of MedDRA include pharmaceuticals, biologics, vaccines and drug-device combination products. MedDRA is available to all for use in the registration, documentation and safety monitoring of medical products both before and after a product has been authorized for sale. Products covered by the scope of MedDRA include pharmaceuticals, biologics, vaccines and drug-device combination products.
- Common Terminology Criteria for Adverse Events (CTCAE, 2010) is a coding for adverse events that occur in the course of cancer therapy. Adverse events is considered any unintended or unfavorable symptom, sign or disease temporarily temporally associated with the use of a medical treatment or procedure that may or may not be considered related to the medical treatment or procedure.
- Foundation Model of Anatomy (Rosse.&Mejino,2008) is an ontology developed and used to describe human anatomy, but as a framework can be applied and extended to all other species. FMA is a reference ontology containing approximately 75,000 classes, over 200,000 terms and hundreds of relationships used to semantically describe human anatomical structures, from microscopic parts, such as macromolecules, cells and their parts, portions of tissues, to macroscopic anatomical structures such as body parts and the whole organism itself. The FMA is available in OWL format and its sub-parts are used for specific anatomy or application ontologies, such as MEDICO, SEMIA and others.
- ICD-10 (International Classification of Diseases Tenth Revision) is issued by the World Health Organization (WHO) as a classification system for diseases and other related aspects in medical practice, including symptoms, abnormal findings, causes of diseases, death certificates and health records (WHO, 1992). The ICD is used by member states of WHO for compilation of national mortality and morbidity statistics, for epidemiological research and for assessment of trends in public health and illnesses. However, ICD-10 is not an ontology in the strict sense, but is usually kept as thesaurus and for classification. There are no relationships defined between the terms, as does not provide distinctions between terms on an ontological level. In recent years, an OWL version of ICD-10 is developed to provide better and more formal way to describe the terms and their relationships at

ontological level. In addition, the next version of ICD is under construction and it is expected to overcome some current limitations.

- RadLex (Radiology Lexicon) is a comprehensive lexicon that aims to provide a unified language for standardized indexing and retrieval of radiological information resources (radLex, 2015). RadLex includes many complex domains that are necessary for radiologists, such as image quality, treatment, anatomic location and uncertainty. RadLex although provides a unified lexicon for radiologist, it is not an ontological frameworks. It has the potential to evolve to an ontology, but further steps are required.

Ontologies for more specific medical domains have been developed as well. For supporting the management and integrations of prostate cancer clinical data, Min et al. developed an ontology for prostate cancer domain (Min et al, 2009). Ontologies for rare diseases like Alzheimer's disease ontology (Malhotra et al., 2013) and other specific health domains continue to be developed and indexed in BioPortal under health category.

Table 1. Feature summary of some representative ontologies in health domain

Name of the ontology	Number of classes	Number of properties	Number of projects	Format
SNOMEDCT	303035	152	21	UMLS
LOINCS	173271	111	5	UMLS
MedDRA	65934	16	4	UMLS
CTCAE	2000000	N/A	1	OWL
FMA	100080	188	15	OWL
ICD10	12451	1	3	UMLS
RADLEX	46059	95	8	OWL

2.2. Semantic Data Repositories

In recent years there is an explosion of health data in terms of heterogeneity and volume. Most of them continue to be published in proprietary formats, and some of them have expressed commitment to publish them in accordance to Semantic Web and with the Linked Data best – practices by exposing data for access via HTTP and putting emphasis on semantically described datasets with possibilities to make relationships and interconnections of this data (see Table 2).

For example, The World Health Organization (www.**who**.int) continues to publish reports on global health problems and also to provide access to enormous statistical data and analyses for important health situations. The WHO's Global Health Observatory (WHOGO, 2015) provides access to over 50 different datasets capturing statistical data and analyzes on environmental health, health systems, HIV/AIDS, immunization, injuries and violence, tuberculosis to name few, which can be categorized by country, region, indicator or topic. A lot of other health institutions such as PubMed (PubMed, 2015), or US National Library of Medicine (NIH, 2015) contain selective links to different huge health database repositories, but most of data continue to be published in proprietary formats, such as PDF or spreadsheets, making difficult their further processing, including the data aggregation and data interconnection process. But, recently we see some serious attempts on utilization of Semantic Web to reduce data processing complexity and resolve some classical integration problems.

CardioSHARE (Vandervalk, McCarthy& Wilkinson, 2008) represents a decentralized web service framework that provides a SPARQL endpoint that enables querying transparently resources in "deep web" from distributed and independent sources. Although the main aim of the project is analysis of clinical data on heart diseases, the framework is generic and can be applied to any type of data. The system is scalable, and more importantly it provides a uniform access to data that is geographically distributed, has diverse ownership and formats, and the generated output is in RDF format.

Zaveri et al. have converted WHOs GHO datasets to RDF using RDF Data Cube Vocabulary (Zaveri et al.,2013). The dataset contains about 3 million triples and it is published as Linked Data using the OntoWiki platform, which provides an SPARQL endpoint for querying the data. Additionally, through OntoWiki it is also possible to browse the data in HTML format.

Bio2RDF (Belleau et al., 2008) is an open-source project that uses Semantic Web technologies to provide one of the largest networks of Linked Data from a diverse set of heterogeneously formatted sources obtained from multiple data providers. The system provides a SQPRQL endpoint that can be used to query around 11 billion triples from 35 different datasets from clinical trials, PubMed and other large data set providers from the biomedical domain.

HCLS Knowledge Base (HCLSKB, 2008) converts, publishes and interlinks biomedical data from 15 distinct data sources from biomedical domain using current Web Technologies. The knowledge base includes data from different sources such as Allain Brain Atlas, PubMed or Medical Subjec Headings (MeSH) to name few.

Other similar projects, such as LinkedCT (Hassanzadeh et al., 2009), or DailyMed (DailyMed, 2015) already have and will continue to produce data from health domain using Semantic Web and Linked Data principles. Datahub (datahub.io) currently indexes hundreds of datasets tagged under health category, and this number probably will have an increased trend in future.

Table 2. Representative Datasets published using Semantic Web

Data source	Number of triples (statements)	Descriptiion
CardioSHARE	N/A	Clinical data on heart diseases and other data in biomedical domain
Bio2RDF	~11 billion	Data from multiple sources
HCLS Knowledge Base	N/A	Data from multiple sources, like from PubMed, clinical trials, etc.
WHOsGHO RDF	~3 million	Data from WHO statistics
LinkedCT	~7 million	ClincialTrials.gov represented in RDF
DailyMed	~100000	Data about marketed drugs

3. Semantic Web in Health – Applications and Interfaces

The projects used to generate semantic data from health domain, such as CardioShare, LinkedCT, Bio2RDF and others already provide capabilities for browsing and searching the underlying semantic data. However, their current interfaces are limited and not intuitive. Those who are not fluent in Semantic Web technologies may have difficulties in searching, browsing, displaying and visualizing the RDF triples. For example, CardiShare interface provides browsing capabilities, a SPARQL endpoint and a query form. However, searching data requires writing queries in SPARQL format as those shown in Figure 1 which requires specific knowledge on underlying data and the SPARQL query language specifics.

```
PREFIX sadi: <http://sadiframework.org/ontologies/properties.owl#>
PREFIX ss: <http://semanticscience.org/resource/>
PREFIX uniprot: <http://lsrn.org/UniProt:>
SELECT ?nameString
WHERE {
    uniprot:P15923 sadi:hasName ?name .
    ?name ss:SIO_000300 ?nameString .
}
```

Fig. 1. Example SPARQL query that can be executed on CardioSHARE

Ding et al. propose a lightweight Semantic Web Portal (SWP) to help users, including those unfamiliar with Semantic Web technologies, to easily publish, browse, display and visualize the semantic data in user – friendly and meaningful way (Ding et al., 2010). The SWP can facilitate the conversion of data from other formats, such as text, relational databases, or spreadsheets into RDF format, in a reasonable period of time and without professional training. Its architectural components allow faceted browsing that can be multi – filtered (based on category, location, etc) that can enable users to explore complex RDF triples in meaningful manner. The system is also able to visualize the data in tile, timeline, Google map and table format. Finally, the search component enables efficient type –based search that can categorize federated RDF triples into different groups based on ontologies. The system was deployed to Indiana University Health Center among others, to look all semantic information in one place. The portal provides access to all information to one patient, such as medication, diagnosis, doctor, diseases, location and time factors.

Tilahun et al. developed Linking Open Data based health information representation, querying, and visualization mashup system based on existing tools (Tilahun et al., 2014). The RDF triples were generated from WHO Global Observatory HIV-related datasets into Fueski SPARQL server. The RDF triples were interlinked with other related datasets, such as DBpedia, Bio2RDF, and LinkedCT using the Silk framework. For users that are not familiar with SPARQL, an interface to search and browse data is provided. And Sgvizler was utilized for the visualization of the semantic data.

Also, a set of web sites and applications that allow viewing the content SNOMED clinical terms have been developed (SCTBROWSERS, 2015). The browsers provide more intuitive displaying and navigation capabilities, flexible search interfaces and different visualization features on SNOMED CT.

4. Data Mining in Healthcare

The large datasets in healthcare are key resource for extraction of hidden patterns. Data Mining brings a set of tools and techniques that can be applied to such data to discover valuable knowledge, such as detection of causes of diseases, identification of medical treatment at lower cost, detection of fraud in health insurance, etc.

This section analyzes the most representative work on applying Data Mining techniques in healthcare data. The analysis is conducted based on Data Mining techniques, namely the classification, clustering and associations rules. The proposed techniques are not specifically designed for semantic health data, but they can be applied to it directly or with small modifications.

Shouman et al. investigates applying K-nearest neighbor(K-NN) classification technique in the diagnosis of heart diseases (Shouman et al., 2012). The approach is applied on Cleveland Heart Disease Dataset, and the experiment was performed with and without voting. It is found that K-NN achieves better accuracy without voting. The experimental results are also used for benchmarking with other classification techniques. Liu et al. proposed an improved fuzzy k-nearest classifier using Particle Swarm Optimization to adapt fuzzy strength parameter neighborhood size for diagnosing thyroid disease (Liu et al., 2012). The authors claim that this approach outperforms other current competitive approaches, Zuo et al. proposed fuzzy k-nearest neighbor for Parkinson disease as well. Chang et al uses an integrated decision tree model to characterize the skin diseases. Decision trees are also used for predicting the survivability of breast cancer patients or those having chronic disease (Chang et al., 2009). Support vector machines(SVM) have also used extensively for classification of diseases. Soliman et al. used SVM for classification of various diseases (Soliman et al., 2010), for analyzing arrhythmia cordis, or breast cancer diagnosis. Neural networks(NN) are proposed for classification of chest diseases, lung cancer, asthma, etc (Er et al., 2010). And Bayesian classification is widely used for analyzing risks that are associated with health (Liu&Lu, 2009), or for classification of psychiatric diseases (Curiac et al., 2009).

Clustering techniques have been used extensively as well. K-means clustering is used to classify the Alzheimer disease into pathologic and non – pathologic groups or to detect the recurrence of breast cancer(Escudero, Zajicek & Ifeachor, 2011). Hierarchical clustering approach is used for grouping the patients according to their length of stay in hospitals (Belcuig, 2009) or for analyzing microarray data. And density based clustering is used in the research work to extract useful patterns from biomedical images (celebi et al., 2005). The approach discovers the area of homogenous colors in images in order to separate unhealthy skin or wounds from healthy skin which is useful for classification and association task.

Association rules are used to detect relationships between symptoms, diseases, drugs and health state. Patil et al. used apriori algorithm for generating association rules that are used to classify whether patients goes to develop type-2 diabetes or not (Patil, Joshi & Toshniwal, 2010). The study indicates that the pre-processing steps are critical, and they propose a modified equal width binning approach to discretizing continuous valued attributes before applying the association rules to the dataset. Abdullah et al. proposed a modification model of apriori algorithm which than is used to extract useful information in medical bill (Abdullah, Ahmad&Ahmed, 2008) . Nahar et al. use association rules for healthy and sick people to discover the factors that cause heart problems in both genders (Nahar et al., 2013). The study conclude that women have less possibility of having coronary heart diseases as compared to man. Ilayaraja et al. also proposed an approach based on apriori algorithm to find frequent diseases in medical data (Ilayaraja& Meyyappan, 2013). Another interesting study of using associations in health is conducted by Noma et al. They propose utilization of pattern frequent tree algorithm in order to discover valuable information from audiometric datasets (Noma,& Ghani, 2012).

5. Limitations, Problems and Challenges

Clinics, health organizations, industry, the science and the human community in general can benefit from the semantic operability of the data, but the survey reveals certain limitations and challenges on the current use of semantic web technologies in healthcare. In order to achieve the expectations of increased productivity from the employment of semantic web technologies in healthcare domain, a number of issues need to be addressed:

(1) The efforts to ontology development in healthcare domain are quite fragmented and non-standardized. For example, several competitive ontologies exist to describe things from the same domain. Reaching common and wide community agreement on ontology development for specific domain is difficult to achieve, but efforts to standardize vocabularies in specific health domains are required.

(2) No single ontology is sufficient to describe the diverse data in health care. The data will continue to be described using different vocabularies. But to integrate such information into a coherent view, various strategies for ontology mapping and alignment are required.

(3)There is an explosion of data, both in volume and diversity. But most of the available datasets in health domain continue to be generated in proprietary formats ranging from those that are convenient, accurate, indexed and well-managed to those that are incomplete, inconsistent, potentially inaccurate and extremely large. There are some serious initiatives to convert such data, but mapping local terms to concepts in standard vocabularies is a complex and resource intensive task. The standard ontologies contain a large number of concepts and relationships, and identifying the correct concept requires specific domain knowledge and knowledge on vocabularies. Resources, tools and strategies are required to automate this process in order to improve the semantic interoperability.

(4) There are few user – friendly interfaces for querying, browsing, displaying and visualizing the semantic health data in meaningful manner. The development of effective interfaces significantly hinders further adoption of the Semantic Web in health.

(5) Data mining brings a set of useful tools and techniques that can be used to discover hidden patterns and knowledge in large healthcare datasets, but:

a) Although the analyzed data mining approaches are not strictly dependent on the format of underlying data, yet few approaches so far are specifically proposed to work with semantic data. The development of data mining tools for semantic data will further increase the usability and usefulness of the Semantic Web.

b) The quality of data is another challenge. Although the current approaches and techniques can be adapted for semantic data, the performance of data mining techniques depends on complete, accurate and relevant data. Therefore, it is very important to generate quality semantic data for training and testing the data mining techniques.

c) Some patterns and result need to be evaluated and used with great care. For example in medical diagnosis, there is a need for data mining techniques with high accuracy, because it is an issue of life or death.

(6) With the data flood, the performance of traditional approaches to data retrieval, data mining and data management needs to be evaluated. What works in small datasets not necessarily works in large scale and complex data as well. Therefore the effectiveness of traditional approaches needs to be proven in large and complex data pools, or new algorithms and tools have to be developed to deal with all of this data.

6. Conclusion

Healthcare sector continuously produces large volumes of heterogeneous and diverse data. However, due to the data integration problem, current information systems are not able to exploit such data in effective and efficient way. As a result of that, retrieval of information and linking data from multiple sources remains difficult task, and much of the information that lies in that data pools remains hidden.

Semantic Web has already gained popularity as platform for data integration and analysis in the health domain. This paper provides an overview of recent developments in this area. The investigation was conducted from three perspectives. First, we reviewed the referential vocabularies and available datasets published according to best principles and practices of Semantic Web. Second, we conducted a survey on human – computer interfaces for querying, displaying and visualizing the semantic health data. Finally, we analyzed the data mining approaches for discovery of hidden knowledge in health data.

Although the value of semantic web technologies in health domain is widely recognized, the survey reveals a number of problems and challenges in current situation. One of the drawbacks is that mappings from proprietary formats to ontology concepts are difficult and intensive task. Maintenance of ontologies and datasets is another challenge. The development of tools that will automate these processes are of great importance. Also the development of use cases, applications and effective interfaces as proof-of –concept will increase the confidence in usefulness of Semantic Web in health domain. Finally, the data mining techniques can be used for extraction of new knowledge, but a number of issues, such as the quality of data are of vital importance to discovery of quality knowledge.

References

HCLSIG. (April, 2015). Semantic Web Health Care and Life Sciences Interest Group.Retreived from http://www.w3.org/blog/hcls/

BioPortal. (April, 2015). The world's most comprehensive repository of biomedical ontologies. Retreived from https://bioportal.bioontology.org/

SNOMEDCT. (April, 2015).Systematized Nomenclature of Medicine - Clinical Terms. Retreived from http://www.ihtsdo.org/snomed-ct

LOINC. (April, 2015). LOINC Logical Observation Identifier Names and Codes. Retreived from http://loinc.org/

MedDRA. Medical Dictionary for Regulatory Activities. Retreived from http://www.meddra.org/

CTCAE. Common Terminology Criteria for Adverse Events (June, 2010). http://evs.nci.nih.gov/ftp1/CTCAE/CTCAE_4.03_2010-06-14_QuickReference_5x7.pdf

Rosse, C., & Mejino Jr, J. L. (2008). The foundational model of anatomy ontology. In Anatomy Ontologies for Bioinformatics (pp. 59-117). Springer London.

WHO. World Health Organization. (1992). The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization. http://www.who.int/classifications/icd/en/

RadLex. Radiology Lexicon. (April, 2015). http://www.rsna.org/RadLex.aspx

Min, H., Manion, F. J., Goralczyk, E., Wong, Y. N., Ross, E., & Beck, J. R. (2009). Integration of prostate cancer clinical data using an ontology. Journal of biomedical informatics, 42(6), 1035-1045.

Malhotra, A., Younesi, E., Gündel, M., Heneka, M., & Hofmann-Apitius, M. (2013). ADO: A disease ontology representing the domain knowledge specific to Alzheimer's disease. *Alzheimer's & Dementia*, *1*, 9.

Whogo. (April, 2015). Global Health Observatory (GHO) data. Retreived from http://www.who.int/gho/en/

PubMed. (April, 2015). US National Library of Medicine, National Institutes of Health. Retreived from http://www.ncbi.nlm.nih.gov/pubmed/

NIH. (April, 2015). U.S. National Library of Medicine. Retreived from https://www.nlm.nih.gov/

Vandervalk, B., McCarthy, L., & Wilkinson, M. (2008). CardioSHARE: Web Services for the Semantic Web. *Semantic Web Challenge*, *2008*.

Zaveri, A., Lehmann, J., Auer, S., Hassan, M. M., Sherif, M. A., & Martin, M. (2013). Publishing and interlinking the global health observatory dataset.*Semantic Web Journal, Special Call for Linked Dataset descriptions*.

Belleau, F., Nolin, M. A., Tourigny, N., Rigault, P., & Morissette, J. (2008). Bio2RDF: towards a mashup to build bioinformatics knowledge systems.*Journal of biomedical informatics*, *41*(5), 706-716.

HCLSKB. (April, 2008). HCLS Knowledgebase. Retreived from http://www.w3.org/TR/2008/WD-hcls-kb-20080404/

Hassanzadeh, O., Kementsietsidis, A., Lim, L., Miller, R. J., & Wang, M. (2009). Linkedct: A linked data space for clinical trials. *arXiv preprint arXiv:0908.0567*.

DailyMed. (April, 2015). DailyMed. Retrieved from http://dailymed.nlm.nih.gov/dailymed/index.cfm

Ding, Y., Sun, Y., Chen, B., Borner, K., Ding, L., Wild, D., ... & Toma, I. (2010). Semantic web portal: a platform for better browsing and visualizing semantic data. In *Active Media Technology* (pp. 448-460). Springer Berlin Heidelberg.

Tilahun, B., Kauppinen, T., Keßler, C., & Fritz, F. (2014). Design and Development of a Linked Open Data-Based Health Information Representation and Visualization System: Potentials and Preliminary Evaluation. *JMIR medical informatics*, *2*(2).

SCTBROWSERS. (April, 2015). SNOMED CT Browsers. Retreived from http://ihtsdo.org/fileadmin/user_upload/doc/browsers/browsers.html

Shouman, M., Turner, T., & Stocker, R. (2012). Applying k-nearest neighbour in diagnosing heart disease patients. *International Journal of Information and Education Technology*, *2*(3), 220-223.

Liu, D. Y., Chen, H. L., Yang, B., Lv, X. E., Li, L. N., & Liu, J. (2012). Design of an enhanced fuzzy k-nearest neighbor classifier based computer aided diagnostic system for thyroid disease. *Journal of medical systems*, *36*(5), 3243-3254.

Zuo, W. L., Wang, Z. Y., Liu, T., & Chen, H. L. (2013). Effective detection of Parkinson's disease using an adaptive fuzzy k-nearest neighbor approach.*Biomedical Signal Processing and Control*, *8*(4), 364-373.

Chang, C. L., & Chen, C. H. (2009). Applying decision tree and neural network to increase quality of dermatologic diagnosis. *Expert Systems with Applications*, *36*(2), 4035-4041.

Khan, M. U., Choi, J. P., Shin, H., & Kim, M. (2008, August). Predicting breast cancer survivability using fuzzy decision trees for personalized healthcare. In*Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE* (pp. 5148-5151). IEEE..

Chien, C., & Pottie, G. J. (2012, August). A universal hybrid decision tree classifier design for human activity classification. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE* (pp. 1065-1068). IEEE.

Soliman, T. H. A., Sewissy, A. A., & AbdelLatif, H. (2010, November). A gene selection approach for classifying diseases based on microarray datasets. In*Computer Technology and Development (ICCTD), 2010 2nd International Conference on* (pp. 626-631). IEEE.

Er, O., Yumusak, N., & Temurtas, F. (2010). Chest diseases diagnosis using artificial neural networks. *Expert Systems with Applications*, *37*(12), 7648-7655.

Liu, K. F., & Lu, C. F. (2009, August). BBN-Based Decision Support for Health Risk Analysis. In *INC, IMS and IDC, 2009. NCM'09. Fifth International Joint Conference on* (pp. 696-702). IEEE.

Curiac, D. I., Vasile, G., Banias, O., Volosencu, C., & Albu, A. (2009, June). Bayesian network model for diagnosis of psychiatric diseases. In *Information Technology Interfaces, 2009. ITI'09. Proceedings of the ITI 2009 31st International Conference on* (pp. 61-66). IEEE.

Escudero, J., Zajicek, J. P., & Ifeachor, E. (2011, August). Early detection and characterization of Alzheimer's disease in clinical scenarios using Bioprofile concepts and K-means. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (pp. 6470-6473). IEEE.

Belciug, S., Gorunescu, F., Salem, A. B., & Gorunescu, M. (2010, November). Clustering-based approach for detecting breast cancer recurrence. In *Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on* (pp. 533-538). IEEE.

Belciug, S. (2009). Patients length of stay grouping using the hierarchical clustering algorithm. *Annals of the University of Craiova-Mathematics and Computer Science Series*, *36*(2), 79-84.

Celebi, M. E., Aslandogan, Y. A., & Bergstresser, P. R. (2005, April). Mining biomedical images with density-based clustering. In *Information Technology: Coding and Computing, 2005. ITCC 2005. International Conference on* (Vol. 1, pp. 163-168). IEEE.

Patil, B. M., Joshi, R. C., & Toshniwal, D. (2010, February). Association rule for classification of type-2 diabetic patients. In *Machine Learning and Computing (ICMLC), 2010 Second International Conference on* (pp. 330-334). IEEE.

Abdullah, U., Ahmad, J., & Ahmed, A. (2008, October). Analysis of effectiveness of apriori algorithm in medical billing data mining. In *Emerging Technologies, 2008. ICET 2008. 4th International Conference on* (pp. 327-331). IEEE.

Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Association rule mining to detect factors which contribute to heart disease in males and females. *Expert Systems with Applications*, *40*(4), 1086-1093.

Ilayaraja, M., & Meyyappan, T. (2013, February). Mining medical data to identify frequent diseases using apriori algorithm. In *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*(pp. 194-199). IEEE.

Noma, N. G., & Abd Ghani, M. K. (2012, December). Discovering pattern in medical audiology data with FP-growth algorithm. In *Biomedical Engineering and Sciences (IECBES), 2012 IEEE EMBS Conference on* (pp. 17-22). IEEE.