

Scientific Report

Designing an Ontology for Knowledge Discovery in Iran's Vaccine

Ghazi Mirsaeed¹, Sh., Abedi Kiasari^{*2}, B., Teimourpour³, B.

1. Department of Information Technology, Razi Vaccine and Serum Research Institute, Karaj, Iran

2. Department of Human Viral Vaccine, Razi Vaccine and Serum Research Institute, Karaj, Iran

3. Department of Information Technology, School of Engineering, Tarbiat Modares University, Tehran, Iran

Received 15 Oct 2013; accepted 07 Dec 2013

ABSTRACT

Ontology is a requirement engineering product and the key to knowledge discovery. It includes the terminology to describe a set of facts, assumptions, and relations with which the detailed meanings of vocabularies among communities can be determined. This is a qualitative content analysis research. This study has made use of ontology for the first time to discover the knowledge of vaccine in Iran. The ontology can be used for knowledge discovery in domain of vaccine in Iran and identification of experts in the field of vaccine, knowledge management, coordination of R & D activities on vaccine production, shaping a culture and flourishing IT in vaccine industry. Visualization, assessment, validation and evaluation of ontology quality, implemented via use of authentic tools and Experts evaluation results obtained from the questionnaires and how knowledge discovery and visualization of knowledge discovery have been discussed. Iran knowledge on vaccinology can be searched with the aid of ontology and can be developed with bipartite Networks of Iran's vaccinology that is result of knowledge discovery. Analysis of the bipartite networks researcher-centers showed 2811 researchers and research centers as nodes and 3341 collaborations cases. Bipartite network researcher - journal analysis showed 2458 researchers and journals as nodes and 3456 collaborations cases

Keywords: Ontology, Vaccine, Visualization, Knowledge discovery, Bipartite Network

INTRODUCTION

Ontology, is an engineered product (Guarino *et al* 1998). Concept is the first step in ontology designing. It is necessary, to specify clearly present concept in thematic domain, meaning relation between them and restriction related to every concept. The created data is significantly for computer. It provides the facility of

conclusion and making it ready for discovering knowledge for computer because the need to the tools for concepts classification, the artificial intelligence's experts in 1990s described with loading the concept of ontology from the field of philosophy. In this way, the concept of ontology is acceded to other domains like Semantic web, linguistics, processing of natural language, Knowledge management, information production. The main merit of ontology is relative

* Author for correspondence. Email: b.abedikiasari@rvsri.ac.ir

freedom than applications. The ontology contains rather public knowledge which can use again with various kind of applications or duties (Noy & MC Guinness 2001). We can present for one domain some different ontology which are all valid. It's clear that the thought of designer's tending to object and his experience in the domain of under studying will have special role in quality of ontology. (Bodenreider & Stevens 2006). In philosophy, clarity between concepts, ontology becomes appear. But computer science is not in this order. Ontology in computer science is said to a set of words and hypotheses which create by considering to the meaning of the word, to describe a special reality use and the reason is increasing the usage of IT (Smith & Welty 2001). Necessity and Background of doing research. In the domain of medic, the challenge of discovering knowledge from various data and biomedical heterogeneous to develop knowledge and concepts among applications. Capable of unity experimental for different goals is very important and complicate. The solidity of data as yet had been as collecting data with the use of maladjusted methods or weak, description for recording data, Storage, Publication and contraction. This problem can use from a common ontology and controlling vocabulary and relation which relate to the definitions. (Zina 2012). Cause, there have been some centers in order to support decision in biomedical activities. Also the possibility of gaining knowledge for everywhere was through the biomedical investigation societies and molecular biologists in advancement of medic understanding (Burgun & Bodenreider 2001). The Background of research in vaccine's ontology the domain of informatics vaccine is coming which the centralization is on the development of applications that uses bioinformatics and immunoinformatics methods and can utilize the work of clinic experts and vaccine institute. In order to access to the vaccine data, from cooperation of vaccine researchers and national centers and international for development programs, the related investigation centers with coordinated vaccine with the

base of violinet (www.violinet.org) data do some activities:

- National center for Biomedical ontology (NCBO)
- Infectious Disease ontology (IDO)
- Ontology biomedical Investigations (OBI)
- Basic Formal ontology (BFO)
- Relation ontology (RO)
- Web ontology language (editing with protégé) (OWL) (Hur et al 2011).

MATERIALS AND METHODS

The usage of vaccine ontology.

- The base of wise on Vaccine Ontology Violinet is the vaccine knowledge base with VO the process of seamless increasment NCBI, Ontobee (www.ontobee.org), etc. with the help of computer reasoning.
- The universality of medical data with VO By using words from different systems of medical ontology, so accretion between vaccine data appears with vaccine ontology.
- Literature on VO Improving the search of vaccine literature research.
- The analysis of vaccine's data. The analysis of clinic's data and vaccine researches apply in combination with other ontology for analyzing general data.
- The analysis of immunity protection network The analysis and prediction of immunity protection networks are achieved according to investigation of data, data analyses, advanced statistics and bioinformatics with the use of vaccine ontology.

The goal of investigation. Discovering the knowledge of Iran vaccine by using entities and defined relations in Iran vaccine ontology the process of subject matters and methods. We specify the method of doing research is some perspective:

Time perspective. Geodesic investigation: The subject is relating to now and by collecting data and information about recent conditions vaccine industry, we will get better and complete understanding which we want to make connected and enrichment of available data in related Center with vaccine.

Goal perspective. Descriptive investigation: The nature, Condition and constitutive elements of our subject are describing without judgment or adjudication.

Collecting data perspective. Open-ended interview: Collecting some parts of information accomplish

through Reference and personal inquiry from experts and specialists of these centers. Review: collecting data achieve through reviewing and checking the available background in libraries or other information centers and searcher motors, also files and informing portals of other countries and famous organizations in vaccine production domain and data sources of goal centers of inside country. Regarding this point that the questions of interview is done by past investigation of review – literature. It has content with specialists.

Society and sample perspective. Collecting data from all terms and activities, is accomplished by relations and related people with investigation society and vaccine production. All information of society is collecting one by one, then census accomplishes. In this field, 107 files of .owl from ontology application sites have analyzed in available bases in web and have collected its review of literature in the Excel file.

The Goal Society.

- The investigation – scientific centers and medical science universities related to Iran vaccine
- The companies' which create vaccine in Iran
- Cyber sites related to vaccine in Iran
- Related articles to vaccine in Iran
- Indicator people in research and production of Iran vaccine
- Vaccine production of country
- Related equipment to create vaccine
- Journals and specialty data bases interior and exterior of country collected subject matters in the form of tables set as following:

1- The table related to the articles which were published the definition of articles related to Iran vaccine ISI (no ISC). It contains some fields like the name, the year of publishing, a journal which was edited the article, the experts who were presented their investigation in the article, the centers which related experts to that center that made the article.

2- The journals' tables introduce the magazines which activated in the field of Iran vaccine. It contains fields like the name of magazine, the efficacy coefficient, the domain, the goal of activity and the address of internet site.

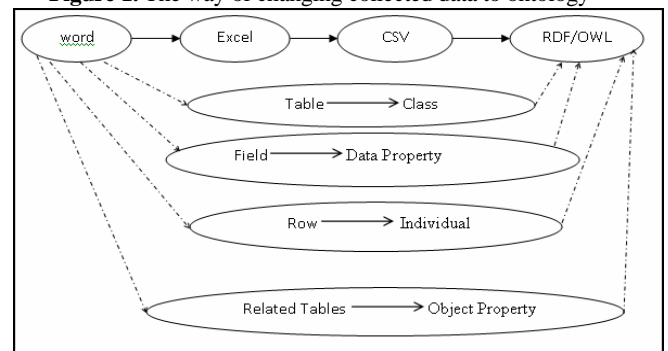
3- The table of centers which activated in the field of vaccine or published articles in the fields of the name of center, the address of internet site.

4- The table of pathogen which contains some fields like name, Disease, the address of page which define pathogen in violent, the research center which is active in the field of pathogen. 5- The table of vaccine contains the special name of product, the trade name of product, the name of main center and international which is active in its field and the name of center which produce vaccine in Iran.

6- The table of equipment contains some fields like the name of Sender Company and supporter.

7- The table of authors contains the name of author and the center of investigation

Figure 1. The way of changing collected data to ontology



Definitions of Ontology. The collected data like the figure 1 are used for definitions of ontology.

Choosing the editing tool and the language of ontology display. protégé software is using as tool to produce, ontology of one amplitude and custom made of data entry. The definitions of entities and scientific relations are specified in protégé in figure 2. This software has capacity if classes definition, classes hierarchy, Variables, restriction of amounts, relation between classes and features of relations. In addition, this tool is available and it's free (Horridge *et al* 2004). The forte of protégé is that, it supports concurrent from software makers, knowledge engineers and experts in different fields.

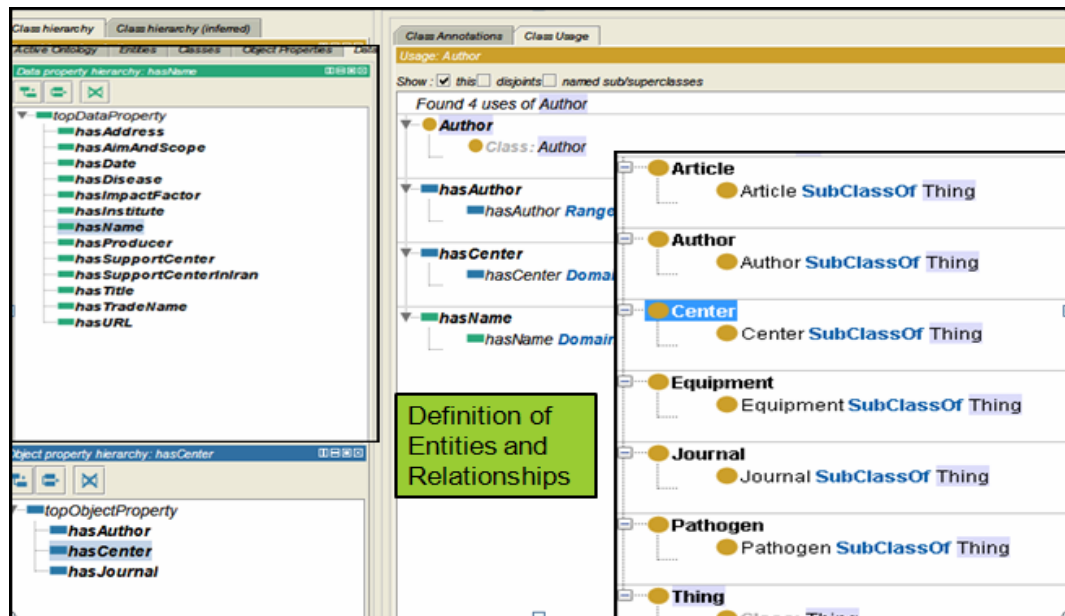


Figure 2. The definition of entities and relations

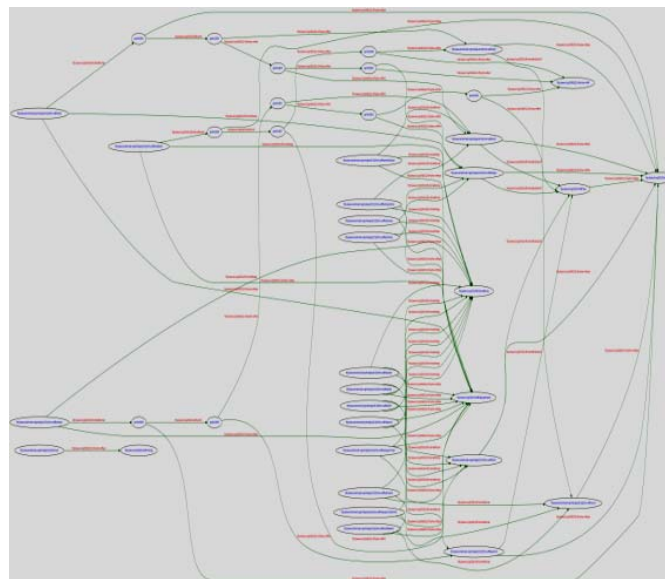


Figure 3. The result Graph from vaccine ontology evaluation

The forte of protégé is that, it supports concurrent from software makers, knowledge engineers and experts in different fields. The most important indicator, Contains the power of reasoning and evaluation of program in declaration of knowledge model and then it is able to future develop with the use of addendum (Mortensen *et al* 2007). Owl is one

the meaning notation which applies clearly with definition of classes, samples and relations in development and creation of ontology. OWL has more facilities for concepts declaration and meaning than XML, RDF and RDFS and because of ability of content display in web is better than other languages (Bechhofer *et al* 2003).

The evaluation of Iran vaccine ontology. The evaluation by using:

- 1- Three online tools ontology validator, its sample is specified in figure 3
- 2- The analysis of ontology structural by using OntoEval tool
- 3- The analysis of adaptation condition of ontology by protégé tool
- 4- The analysis of available ontology entities conclusion by using protégé tool
- 5- The validity of ontology with valid tool Jena Eyeball
- 6- Presenting questionnaire to experts and qualitative analyzing responses presenting questionnaire to experts for evaluation of the certification's rate of available ontology competence designing questionnaire, is done for evaluation of the certification's rate of available ontology by expert person. (TANKELEVI & DAMAŠEVI & CIUS 2009). The rate of Cronbach's alpha for all the questionnaires is 0.798 and the edited questionnaire is capable of less evaluation. The available ontology has evaluated by experts with questionnaire distribution between 30 people from vaccine experts which there are 20 completed answer sheets the experts, qualitative evaluated ontology in the form of completeness, Consistency, Conciseness, Preciseness and Clarity. The applied questions were considered by experts and specialists related to Iran vaccine and the available ontology were answered to them, is as following:
 - what are the important entities in the domain of vaccine?
 - Which is the prophylaxis disease through vaccine in Iran?
 - What are the main production and disposability vaccines in Iran?
 - Which are the active journals related to Iran vaccine?
 - Which valid articles related to Iran vaccine, have ever published?
 - Which are the list of researchers and experts related to Iran vaccine?
 - Which are the equipment in vaccine industry?

The experts' suggestion: Separately.

- 1- The experts social network analyses separately with complete details in tools
 - 2- It's better, in the time of developing ontology considers the interest of scientists which work under different fields of vaccine science.
 - 3- If the argument of vaccine ontology results in the analyzation of data in Iran vaccine, So it is very suitable and has more importance.
 - 4- It's a suitable tool and able to expand. For example we can Contemplate it under the domain of vaccines qualitative tests in the phase of ontology development.
 - Discovering the knowledge of Iran vaccine by checking these three networks and the various available relations in it, we can get to the vaccine knowledge of discovery in domain of Science and discovery the available relations.
 - Co – Authorship Network
 - Bipartite Network from Author and organization
 - Bipartite Network from Author and Journal
- Finding authorship with the most published articles and the Cooperation relation between them we change the file of created ontology in protégé with OWL format to BibTex. Then we use it in SCI2 and GEPHI finding the relations in network. We analyses Bipartite Network from Author and organization we review different relations in it: 1- Center – Authors (Direct Relation) 2- Centers– Author (Direct Relation) 3- Center – Center (Indirect Relation) 4- Authors – Authors (Indirect Relation) have the most co operation in the field of Iran vaccine according to table 1 and Figure 4. With accuracy in getting outcomes from the illustrator of Iran vaccine knowledge with centers parameters and scientists have shown which is the top relation between nodes and in other words the investigation centers which are active, in term of Iran vaccine knowledge in the field of science is, Pasteur institute and Razi institute and Tehran medical science university and Tarbiat Modares university were the activest. Then we analyses Bipartite Network from Author and journal according to Figure 5, 6 and table 2. At first we review different relations:
- 1- Journal – Authors (Direct Relation)

- 2- Journals – Author (Direct Relation)
- 3- Journal – Journal (Indirect Relation)
- 4- Authors – Authors (Indirect Relation)

The analysis in present graph and in getting results from the knowledge illustrator of Iran vaccine with journals parameters and scientists (authors) showed that vaccine journal has 198 connecting crest. The Impact Factor of this journal is 3.766. So Iran vaccine experts had the highest connection with the best journal in the field of vaccine.

The results of Iran vaccine ontology.

- Using IT in vaccine science and industry
- Knowledge management



Figure 4. The relation network for Research Centers

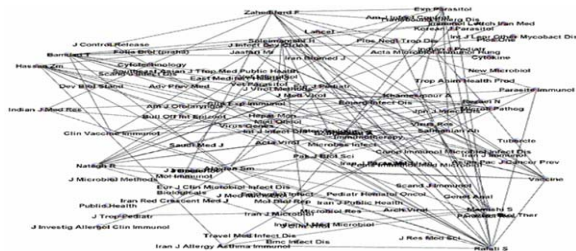


Figure 5. The relation network vaccine of Journals – Author

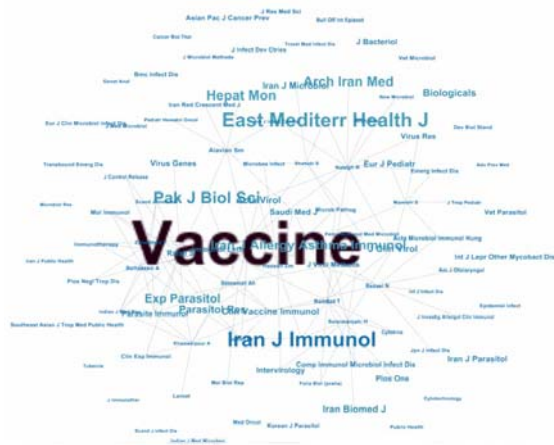


Figure 6. The relation network of journals

Table 1. The Centers which have the most cooperation in the field of vaccine

Research center	Cooperation
Tehran University of Medical Sciences	340
Pasteur Institute	304
Razi Vaccine Serum Research Institute	225
Tarbiat Modares University	119
Shaheed Beheshti University of Medical Sciences	55
Mashhad University of Medical Sciences	49
Shiraz University of Medical Sciences	44
Safed University of Applied Sciences	33
Iran University Of Medical Sciences	27
Kermanshah University of Medical Sciences	26
Baghiatallah University of Medical Sciences	24

Table 2. The journals which have the most cooperation in the field of vaccine

Journal Name	Cooperation
Vaccine	198
Iran J Immunol	90
East Mediterr Health J	85
Pak J Biol Sci	72
Hepat Mon	61
Arch Iran Med	60
Iran J Allergy Asthma Immunol	58
Exp Parasitol	53
Parasitol Res	46
Biologicals	43

- Coordination facilities and scientific relations
- Making capacity, increasing the field of using researches results
- Quantitive and qualitive development in the field of country vaccine

- Improving investigation culture in beneficiary specially users and diplomats

The Identification of experts of this field The themes and subjects getting from research.

1- Checking articles and journals related to Iran vaccine in ISC and adding it to available data in this investigation and designing the knowledge outwork in branch of vaccine science

2- Checking all effects parameters in technical and innovation science of Iran vaccine, illustrator and display the relations and discovering knowledge related to them

3- The way of correlation of Iran vaccine experts with available ontology

4- Making accrete ontology with available and past system onto knowledge

5- Ontology management in semantic web

6- Notional arguments in the field of vaccine ontology development (Science, technical and innovation)

7- Ontology security

8- Ontology correlation of Iran vaccine with other presented ontology in websites like violinet.

RESULTS AND DISCUSSION

Ontology is a new way for investigators and experts of country in order to access and connect quickly to society information and Iran vaccine industry. Available gaps with the help of modifications, remove vaccine ontology concepts coordination in researches activities and production in the field of vaccine, facilities display and available scientific potential, the consolidation of cooperation and the scientific relation of making accrete, increasing the field of using researches results, improving culture of investigation in beneficiary specially users and diplomats, improving qualitative and quantities researches related to vaccine's field of country is the matter which this investigation for the first time in our country generally is set. By making ontology visualization in vaccine research and product, provides the ability to access to vaccine researches, the experts of this field and related Research center for Iran vaccine science.

Conflict of Interest

The authors have no conflict of interest.

References

- Bechhofer, S., Harmelen, F., Hendler J., Horrocks, I., Deborah McGuinness, L., Patel Schneider, P. and Stein, L. (2004). OWL Web Ontology Language W3C Proposed Recommendation.
- Burgun, A. and Bodenreider, O. (2001). Mapping the UMLS Semantic Network into General Ontologies. *International Journal*, Pp. 86-90.
- Guarino, N. (1998). Formal Ontology and Information Systems. *Proceedings of FOIS' 98*, Trento, Italy, 6-8 June 1998. Amsterdam, IOS Press, Pp. 3-15.
- Horridge, M., Knoblauch H., Rector A. and Stevens, R. (2004). A Practical Guide To Building OWL Anthologies Using The Protege-OWL Plugin and CO-ODE Tools, Edition 1.0, *The University Of Manchester And Stanford University*.
- Hur, J., Xiang, Z., Feldman, E. L. and He, Y. (2011). Ontology-based *Brucella* vaccine literature indexing and systematic analysis of gene-vaccine association network. *BMC Immunol Journal*. 2011 Aug 26;12:49. doi: 10.1186/1471-2172-12-49.
- Mortensen, J. M., Horridge, M., Musen, M. A. and Noy, N. F. (2012). Modest Use of Ontology Design Patterns in a Repository of Biomedical Ontologies. Stanford Center for Biomedical Informatics Research Stanford University, Stanford CA 94305, USA.
- Smith, B. and Welty, C. (2001). Ontology: Towards a New Synthesis. *Proceedings of the international conference on Formal Ontology in Information Systems* Volume 2001, Pp 3-9.
- TANKELEVIĆ CIEN'E, L. and DAMAŠEVIĆ CIUS, R. (2008). Characteristics of Domain Ontologies for Web Based Learning and their Application for Quality Evaluation. *Informatics in Education*, 2009, Vol. 8, No. 1, 131-152.
- Zina, N. (2012). Development of Ontology for the Representation of Adverse Drug Events of Diabetes Disease. *International Journal*, Vol. 42 No. 20, Pp. 10-16.

