Egyptian Computer Science Journal (ISSN-1110-2586)
Volume 41– Issue 2, May 2017

Developing a System for Medical Ontology Evolution

Mariam Gawich, Marco Alfonse, Mostafa Aref, and Abdel-Badeeh M. Salem
Faculty of Computer and Information Science. Ain Shams University. Cairo, Egypt
Email: mariamjawich@gmail.com, marco@fcis.asu.edu.eg, mostafa.aref@fcis.asu.edu.eg, abmsalem@cis.asu.edu.eg

Abstract

Ontology evolution is defined as the process of updating the ontology according to the changes in the domain. Current ontology evolution techniques in the medical domain focus on the consistency after the ontology evolution and ignore it during the evolution process. This paper presents a novel evolution system that takes into account the ontology consistency. It relies on the use of standard medical resources that reflect the changes that are occurred in the medical domain. Moreover, the system makes use of a database that contains the scientific and Egyptian commercial names of medicine used in autoimmune diseases.

Keywords: *Knowledge engineering, Ontology engineering, Ontology evolution, Ontology enrichment, Ontology population*

1. Introduction

Various researchers have defined the notion of ontology evolution. Haase and Stojanovic have defined it as “the process of adapting and changing the ontology in a consistent way”. Plessers [1] has defined the ontology evolution as “the process of adaptation of ontology to arisen changes in the corresponding domain while maintaining both the consistency of the ontology itself as well as the consistency of depending artifacts. Examples of depending artifacts include other ontologies, websites, web applications, etc. which depend on the ontology”.

Fouad Zablith [2] proposed an ontology evolution cycle which is composed of five tasks; detecting the need for evolution, suggesting changes, validating changes, assessing impact and managing changes. Detecting the need for evolution can be decided by the user or the data sources that are relevant to the ontology. Suggested changes can be presented by the structured and unstructured data sources. Validating changes can be established by the use of reliable data sources relevant to the domain of interest or by using formal properties and constraints to ensure the consistency of the ontology. The assessing impact task is executed through the measure of ontology changes impact on the artifact that deals with the evolved ontology. Managing changes can be executed through the recording of changes that are applied to the ontology and the use of change detection tools to manage ontology versions.

M.Ben Messaoud and colleagues proposed an approach [3] for biomedical ontology evolution (SemCado) that is based on the use of Causal Bayesian Network (CBN). The approach maps the nodes presented by the CBN to the correspondent concepts in the ontology. The SemCado reuses the knowledge provided by the CBN to evolve the ontology. The approach extends the MyCado algorithm [4,5] that builds the CBN from the observational data set. In addition, SemCado uses the semantic distance calculus [6] to extract the most informative semantic causal relations from the CBN. The semantic causal relations will be used to evolve the ontology after their validation by experts. The outputs involve CBN and evolved ontology.

Julio Cesar and colleagues proposed DyKOSMap [7] to evolve the mapping between two evolved Knowledge Organizations Systems (KOS). The KOS involves biomedical classifications, ontologies, semantic networks, terminologies and taxonomies. DyKOSMap inputs are current KOS, evolved KOS and current mappings. The approach relies on the use of change patterns proposed by Hartung [8] and heuristic rules. The change patterns are used to detect complex changes that are occurred on KOS and heuristic rules represent the addition, removal and modification actions. The output of this approach is an evolved mapping that is consistent to the evolved KOS.

Anny Kartika and colleagues [9] proposed an approach for the evolution of a sub ontology used in a distributed health care enterprise. The inputs involve a health base ontology and its change log. The approach relies on the identification of semantic operations that are occurred on the base ontology. The semantic change operations are derived from the comparison between the base ontology and the evolved base ontology using the change log. The output is an evolved sub ontology that is coherent with the evolved base ontology.

Siqiang Tao and colleagues proposed the mining relations reversals approach [10] to discover the transitive closure and the reversal relations that can be occurred between SNOMED [11] versions. It applies the MapReduce algorithm [12] which takes concepts and is-a relations provided by two SNOMED versions as input. The MapReduce algorithm detects the transitive closure pairs and the reversal relations between concepts that are provided by two SNOMED ontologies. The user can visualize the contradiction of changes that were occurred during the ontology evolution.

Kristina Harris and colleagues proposed the Semi Automated Ontology Management (SEAM) [13]. Its inputs involve clinical and biomedical texts. The system objective is the extraction of terms, synonyms and relationships concerning a particular disease. For term extraction, SEAM applies tokenization, chunking, term, frequencies, C-value [14] and termhood [15] algorithms. It uses the UMLS metathesaurus [16] to get the synonyms of terms as well as the related terms. For the relationships extraction, SEAM applies lexico syntactic pattern matching algorithm [17] to get the synonym relationships located in the clinical and biomedical texts. The SEAM output involves a set of recommended terms, synonyms and relationships that can be used for medical ontology evolution.

The objective of this paper is the presentation of a new system for ontology evolution in medical domain that takes into account the ontology consistency. This paper is organized as follows; section 2 presents the proposed system architecture, section 3 provides the evaluation of the proposed system and section 4 contains the conclusion and future work.

2. The Proposed System Architecture

As figure 1 demonstrates, the proposed system consists of three components: The medical data sources (medical database, UMLS and RxTerms), the recommendation of scientific name and the ontology evolution process. The system enables the user (physician, patient, heath care personnel … etc.) to apply the basic changes and complex changes on the ontology. The user can apply the basic changes through the addition or deletion of classes, instances, data properties and object properties. The system checks the data sources to make sure that the medical term, which the user wants to add, exists in the medical domain. The medical term involves treatment name and clinical finding. The scientific medical term can have several variants that can be expressed in an abbreviated form or trade name, for example the variant of C-reactive protein is CR and the variants of hydroxychloroquine treatments are hydroquine and plaquenil. Concerning the complex changes, the system enables the user to apply directly the merge, split and move operations on the ontology.

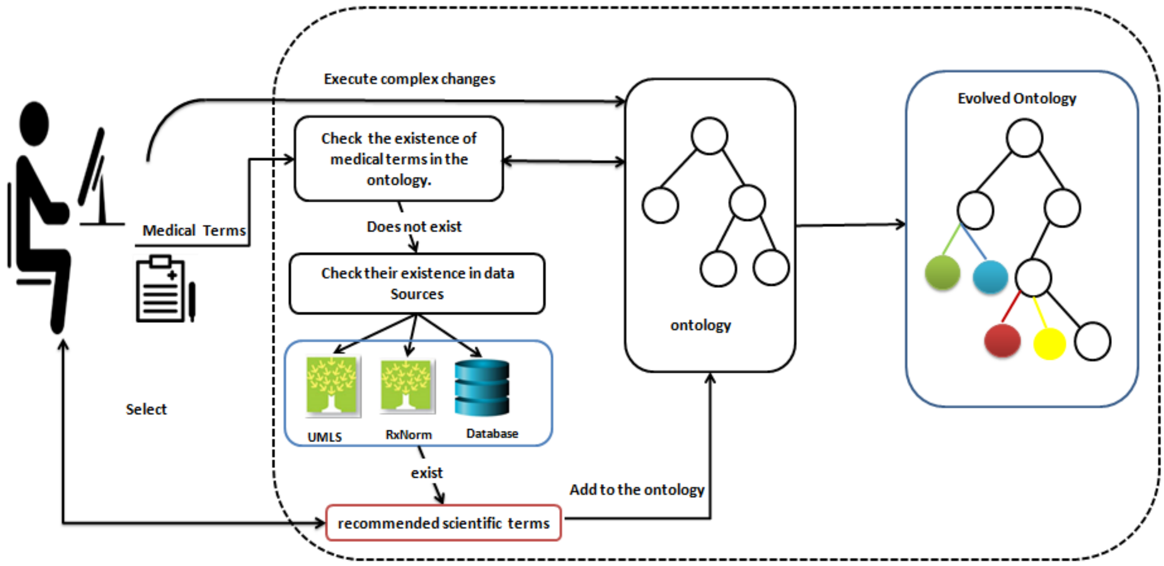

Figure 1. The Proposed Ontology Evolution System

2.1 Medical Data sources

A. Treatment Database

The database contains the scientific name of the treatment and its commercial names used in Egypt (local commercial names). The data are collected from Egyptian Drug Authority [18] and the ATLAS Egyptian Drugs database [19]. The treatment database is used in the system to get the scientific name of the Egyptian commercial treatment name.

B. UMLS metathesaurus

The UMLS stands for Unified Medical Language System. It was created by the national library of medicine and it has three main components; the semantic browser, the metathesaurus and the specialist lexicon. The UMLS metathesaurus involves several data files [20], one of them is called MRCONSO.RRF which contains concepts and their equivalent names. Each concept in the UMLS metathesaurus is attached to a semantic type. One of the semantic types is findings that include sub semantic groups: laboratory/test result and sign/ symptom. The version of the UMLS metathesaurus used in this system is 2016AB; this version contains 9417453 concept names [21] that are used in biomedical and medical domain. The UMLS metathesaurus is used in the system to ensure the existence of the medical term inserted by the user and to detect the scientific name of the term.

C. RxTerms

It is a terminology interface [22] that is based on the RxNorm [23] which stands for standardized nomenclature for clinical drugs. It was developed by the national library of medicine. It contains the normalized drug name, full generic name as well as the semantic branded drug name. The RxTerms files can be downloaded and imported in MySQL database. The RxTerms is used in the system to get the treatment that doesn't exist in the UMLS metathesaurus.

2.2 Recommendation of scientific names

The user enters a medical term then this component will check the ontology to make sure that the medical term does not already exist. If the medical term is a commercial treatment name that is used only in Egypt, the component will verify that it doesn't exist in the ontology then it will check the database to get the scientific name of the treatment and rechecks the ontology to make sure that the treatment does not exist with its scientific form. The component will display the recommended scientific name to the user who will make a decision either to add it in the ontology as an instance attached to a specific class or creating a new class and attach the recommended scientific treatment name as instance. In case that the treatment is not in the database, the component will check the UMLS to get the scientific name of the treatment taking into account the manipulation of composed terms. The component will recheck the ontology to figure out the existence of the scientific term provided by the UMLS, if the component detects its existence in the ontology, a message will be displayed to the user that mentions the existence of the equivalent term with its scientific term.

In case that the UMLS doesn't contain the scientific name of the treatment, the component will connect to the RxTerms to get it. Moreover, the component will return to the ontology to verify the existence of the scientific term provided by the RxTerms. The component will display to the user the recommended scientific treatment name if it doesn't exist in the ontology.

In case that the medical term isn't a treatment name, the component will connect to the UMLS to get its scientific name. For example:

- If the user enters a laboratory test such as CBC, the UMLS will detect its scientific name which is Complete Blood Count.
- If the user enters a sign such as swelling joint, the UMLS will detect its proper name which is Joint Swelling.

2.3 Ontology Evolution Process

The ontology evolution component takes into consideration the basic and complex changes that can be applied on the ontology.

A. Basic Changes

The recommended scientific term can be added to the ontology as instance. The component will display the existing classes to the user who can choose the suitable class that the instance will be attached to. Moreover, the component enables the user to create a new class and attach the recommended scientific term to it. The system proposes to the user to execute the other basic operations such as add/delete a class, add/delete an individual, add/delete an object property, add/delete an object property assertion and add/delete a data property.

B. Complex Changes

The ontology evolution component enables the user to execute the merge, split and move operations. For the merge operation, the component asks the user to insert the two classes then the component checks their existence in the ontology and asks the user to enter the class name in which the two classes will be merged; the user can create a new class or choose one of the classes that already exist in the ontology. The component takes into account the subclasses, individuals, data properties and object properties that are attached to the classes which will be merged. For the split operation, the component enables the user to split an existing class into two new created classes. For the move operation, the component enables the user to change the superclass of a given class (change class hierarchy). The component takes into account the subclasses, individuals, data properties and object properties that are attached to the classes which will be used in the split or the move operation.

3. The Proposed System Evaluation

The ontology evolution techniques apply the evaluation in several ways; comparison of the technique with others, calculation of precision and recall, and the intervention of human expert or ontology engineer. The SemCado executes the evaluation through its comparison with MyCado algorithm using the same observational data which are gene expression [24] and gene ontology [25]. The evaluation of DyKOSMap relies on the calculation of precision, recall and f-measure [9]. The subontology evolution approach is evaluated by its comparison with the re-extracted subontology from the evolved main ontology, the evaluation takes into account the content and the number of implemented change operations on concepts and relations. The SEAM evaluation relies on the intervention of human experts. The evaluation of the proposed system is based on the tasks of the ontology evolution process proposed by Zablith and also its comparison with the other techniques.

3.1 System Evaluation according to ontology evolution cycle

Table 1 demonstrates the applied ontology evolution tasks in the proposed system and in the other approaches.

Table 1.System Evaluation according to ontology evolution cycle

Ontology Evolution Techniques	**Detecting the need for the evolution**	**Suggesting Changes**	**Validating Changes**	**Assessing Impact**	**Managing changes**
SemCado Approach	Yes	Yes	Yes	No	No
DyKOSMAP Approach	Yes	Yes	Yes	No	Yes
Sub Ontology Evolution Approach	No	Yes	Yes	Yes	No
Mining Relation Reversals Approach	Yes	No	No	No	No
SEAM	Yes	Yes	Yes	No	No
The proposed System	Yes	Yes	Yes	No	Yes

All of the ontology evolution techniques focus on detecting the need for evolution, suggesting change and validating changes tasks. Only the sub ontology evolution approach executes the assessing impact task. Both of the DyKOSMAP and the proposed system perform the managing changes task. The mining relation reversals approach focuses on the discovery of inconsistencies that can be occurred between old SNOMED ontology version and evolved SNOMED ontology, therefore it doesn't execute most of the ontology evolution tasks.

For the proposed system, the task of detecting the need for evolution relies on the user and data sources. The user will determine the need to update the ontology. Moreover, the system enables the communication with the updated structured data sources (the UMLS, the RxTerms and the database) that contain the relevant knowledge.

For the suggestion changes task, the candidate changes are provided by the structured data sources. They provide the updated trade name and the scientific names of a treatment.

Concerning validating changes task, the validation of change relies on the data sources to find the medical term which the user searches before adding it to the ontology. Moreover, the validation is implemented through the use of a set of rules that prevents the changes that may cause inconsistency. For example:

- The system will prevent the user to add a class, an individual, an object property or a data property that is already existed in the ontology.
- If the user decides to apply merge/split operations, the system executes a set of rules that ensures that the class, which will be merged/ splitted, keeps its children after the merge or the split operation.

For the assessing impact task, the proposed system can be integrated with a health care application in order to assess the impact of ontology changes on this application.

For the managing changes task, the system applies the changes to the ontology and keeps these changes in a log file.

3.2 Comparison with other systems

The comparison between the proposed system and the other ontology evolution approaches is based on the following points

A. Consistency verification [26]

Does the approach incorporate the application of rules or constraints to prevent the changes that can violate the ontology consistency?

B. Change implementation [27]

Does the approach allow the user to accept or decline the proposed changes? In addition, does the approach show the consequences of the change made on the ontology application?

C. Kind of changes

Does the approach execute the basic and complex changes?

D. Relation discovery [28]

Does the use of data sources present the ancestor of a concept?

Table 2 shows the comparison between the proposed system and the other medical ontology evolution approaches.

Table 2. Comparison of the Proposed System with the other Medical Ontology Evolution Approaches.

Approach Name	Input	Consistency Verification	Change implementation	Kind of changes	Relation discovery	Output
SemCado Approach	Observational dataset and single domain ontology	Yes	Yes	Basic	Yes	Causal Bayesian network with suggested new causal relations and enriched ontology
DyKOSMAP Approach	New published versions on KOS and current mapping	Yes	No	Complex	Yes	Up to date mapping between evolved KOS
Sub Ontology Evolution Approach	Health ontology and its change log	Yes	Yes	Basic and Complex	Yes	Creation of evolved sub ontologies for every health applications
Mining Relation Reversals Approach	For transitive closure, Ontology concept and isa relations. For inversal relations, Transitive closures' concept pairs of two versions	Yes	No	detects the reversal relations	Yes	Transitive closure and inversal relations between two ontology versions
SEAM	Clinical and biomedical text	Yes	No	detects changes in the domain	Yes	Recommended terms, synonyms and relationships
The Proposed System	Medical ontology and Medical term/ Complex changes	Yes	Yes	Basic	Yes	Recommended standardized terms to be added in the ontology. Evolved ontology

The SemCado approach; its inputs are observational dataset and domain ontology. The consistency verification is executed by the use of semantic distance calculus that detects the most informational relations. In addition, the SemCado applies a set of rules to ensure the ontology consistency and also the experts validate the discovered causal relations. Concerning the change implementation, the SemCado enables the user to accept or decline the changes. The SemCado applies the basic changes on the ontology. It uses the causal bayesian network that detects the relation path of a concept. The output of the SemCado involves a causal Bayesian network and new discovered causal relations that can be used to evolve the ontology.

For the DyKosMap, its inputs are new published KOS versions and current KOS mapping. The consistency verification is realized by the use of heuristic rules to get the evolved mapping consistent with the evolved KOS. Since the change patterns are specific for KOS, the evolution mapping relies on the medical domain. The DyKosMap executes complex changes and its data sources can induce the relations between concepts. The output is an up to date mapping between evolved KOS.

For the sub ontology evolution approach, its inputs are base ontology file and its change log file. The consistency verification is realized by the application of rules to extract the sub ontology that was affected by the evolution of the whole ontology. For the change implementation, the approach displays the candidate terms to the user who can accept or decline them. The approach executes basic and complex changes. Since the evolution of the sub ontology approach relies on the evolved whole ontology, the relations between concepts can be detected from the whole ontology. The output involves a set of evolved sub ontologies that are suitable for each health application.

Concerning the mining reversals relations, it has two pairs of input; first pair consists of concept node and is-a relations and the second pair involves the transitive closures' concept pairs of two ontology versions. The approach applies the MapReduce algorithm that determines reversal relations between concepts that are provided by two SNOMED versions. The approach enables the user to view the changes that are occurred on SNOMED versions but the change implementation doesn't executed. Since the mining reversal relations approach uses the MapReduce algorithm, it can discover the relations between concepts. The outputs of this approach are the transitive closures and the reversal relations.

Concerning the SEAM approach, its inputs are clinical and biomedical text. The consistency verification is executed by the use of term frequency, lexico syntactic patterns and term filter. It recommends only the synonym terms and relationships. Since the SEAM approach uses the UMLS as a knowledge source, the ancestor of the concept can be detected. The outputs of SEAM are recommended terms, recommended synonyms and recommended relationships to be used for change operations in the ontology evolution.

For the proposed system; its inputs are the ontology to be evolved and the medical terms/complex changes. The need for the evolution depends on the structured data sources as well as the user. The proposed system uses a set of rules for each kind of change to ensure the consistency of the evolved ontology. The use of data sources suggest the recommended standardized term to the user who decides either to add it to the ontology or not. Although the system uses medical structured data sources (the UMLS, the Rxterms and the drugs database), it can be used with any domain ontology if it is connected to the structured data sources of the domain of interest. The proposed system enables the user to implement the basic and complex changes.

4. Conclusion and Future Work

Although both of the DyKOSMap and the proposed system execute the ontology evolution tasks except the assessing impact, the DyKOSMap focus on the evolution of mapping instead of the ontology evolution. The proposed system applies both of the basic and the complex changes taking into account the ontology consistency. Although the proposed system deals with medical data sources, it can be applied in other domains using the relevant data sources. The use of standard data sources, such as UMLS and RxNorm, ensures the validation of the medical term that will be inserted in the ontology by the user. The use of drugs database in the system enables the user to insert the treatment with its common localized term used in Egypt. The use of the scientific names of medicine in the ontology allows the integration of the evolved medical ontology with other medical ontologies. In future work, the proposed system can be integrated in a distributed health care enterprise or in a clinical decision support system which enables the assessing impact task to be implemented.

References

[1]. P. Plessers, O. De Troyer and S. Casteleyn, ”Understanding ontology evolution: A change detection approach”, Web Semantics: Science, Services and Agents on the World Wide Web, vol. 5, no. 1, pp. 39-49, 2007.

[2]. F. Zablith, G. Antoniou, M. d'Aquin, G. Flouris, H. Kondylakis, E. Motta, D. Plexousakis and M. Sabou, "Ontology evolution: a process-centric survey", The Knowledge Engineering Review, vol. 30, no. 01, pp. 45-75, 2015.

[3]. M. Ben Messaoud, P. Leray and N. Ben Amor, "SemCaDo: A serendipitous strategy for causal discovery and ontology evolution", Knowledge-Based Systems, vol. 76, pp. 79-95, 2015.

[4]. S. Meganck, “Towards an Integral Approach for Modelling Causality”, Ph.D. Thesis, INSA Rouen and Vrije Universiteit Brussels ,2008.

[5]. S. Meganck, P. Leray and B. Manderick,”Learning causal bayesian networks from observations and experiments: A decision theoretic approach”. In International Conference on Modeling Decisions for Artificial Intelligence ,Springer Berlin Heidelberg, pp.58-69, 2006.

[6]. E. Blanchard, M. Harzallah, H. Briand and P. Kuntz, “A typology of ontology-based semantic measures”, 2nd INTEROP-EMOI Open Workshop on Enterprise Models and Ontologies for Interoperability at the 17th Conference on Advanced Information Systems Engineering (CAISE’05), vol. 160, pp.407–412, 2005.

[7]. J. Dos Reis, C. Pruski, M. Da Silveira and C. Reynaud-Delaître.”DyKOSMap: A framework for mapping adaptation between biomedical knowledge organization systems”, Journal of Biomedical Informatics, vol. 55, pp. 153-173, 2015.

[8]. M. Hartung, A. Groß and E. Rahm, "COnto–Diff: generation of complex evolution mappings for life science ontologies", Journal of Biomedical Informatics, vol. 46, no. 1, pp. 15-32, 2013.

[9]. A. Sari, W. Rahayu and M. Bhatt, "An approach for sub-ontology evolution in a distributed health care enterprise", Information Systems, vol. 38, no. 5, pp. 727-744, 2013.

[10]. S.Tao, L.Cui,W. Zhu, M.Sun, O.Bodenreider and G-Q.Zhang. “Mining Relation Reversals in the Evolution of SNOMED CT Using MapReduce”. AMIA Summits on Translational Science Proceedings, pp.46-50, 2015.

[11]. https://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html.[Accessed: 25-March-2017].

[12]. J.Dean and S. Ghemawat. "MapReduce: simplified data processing on large clusters". Communications of the ACM, vol.51,no.1, pp.107-113, 2008.

[13]. K. Doing-Harris, Y. Livnat and S. Meystre."Automated concept and relationship extraction for the semi-automated ontology management (SEAM) system", Journal of Biomedical Semantics, vol. 6, no. 1, 2015.

[14]. K.T. Frantzi, S.Ananiadou and J.Tsujii. "The c-value/nc-value method of automatic recognition for multi-word terms". Research and Advanced Technology for Digital Libraries. Berlin Heidelberg: Springer, pp. 558–604, 1998.

[15]. Q. Zeng, T. Tse, G. Divita, A. Keselman, J. Crowell, A. Browne, S. Goryachev and L. Ngo, "Term Identification Methods for Consumer Health Vocabulary Development", Journal of Medical Internet Research, vol. 9, no. 1, p. e4, 2007.

[16]. UMLS® Reference Manual [Internet]. Bethesda (MD): National Library of Medicine (US); 2009 Sep. 4, Metathesaurus - Original Release Format (ORF).Available: http://www.ncbi.nlm.nih.gov/books/NBK9682/. [Accessed: 30-April-2017].

[17]. K. Liu, W. Chapman, G. Savova, C. Chute, N. Sioutos and R. Crowley, "Effectiveness of Lexico-syntactic Pattern Matching for Ontology Enrichment with Clinical Documents", Methods of Information in Medicine, vol. 50, no. 5, pp. 397-407, 2010.

[18]. http://www.eda.mohp.gov.eg/SearchRegDrugs.aspx. [Accessed: 30- March- 2017].

[19]. https://apkpure.com/atlas-of-the-egyptian-drugs/com.RapidDevelop.Atlas. [Accessed: 26-April- 2017].

[20]. https://www.ncbi.nlm.nih.gov/books/NBK9685/. [Accessed: 27- April- 2017].

[21]. https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/release/statistics.html. [Accessed: 30- April- 2017].

[22]. https://rxterms.nlm.nih.gov/ [Accessed: 31 –March-2017]

[23]. MD Clement McDonald, BE Bray. "RxTerms - a drug interface terminology derived from RxNorm". AMIA 2008 Symposium Proceedings, pp 227–231, 2008.

[24]. PT. Spellman, Gavin Sherlock, M.Q. Zhang, V.R. Iyer, K.Anders, M.B. Eisen, P.O. Brown, D.Botstein, and B.Futcher, " Comprehensive Identification of Cell Cycle–regulated Genes of the Yeast Saccharomyces cerevisiae by Microarray Hybridization", Molecular Biology of the Cell. 9, 3273–3297,1998.

[25]. http://geneontology.org/. [Accessed: 30- April- 2017].

[26]. P.Plessers and O.De Troyer. "Ontology change detection using a version log". Proceedings of the 4th International Semantic Web Conference (ISWC- 05), PP. 578-592, 2005.

[27]. L.Stojanovic."Methods and tools for ontology evolution". PhD thesis, University of Karlsruhe, 2004.

[28]. F. Zablith. "Evolva: A Comprehensive Approach to Ontology Evolution", Lecture Notes in Computer Science, pp. 944-948, 2009.