

Ontology Evolution Approaches in Medical Domain

Mariam Gawich, Marco Alfonse, Mostafa Aref, and Abdel-Badeeh M. Salem

Abstract— The growth of medical applications that work with ontologies in health enterprises creates the need to ensure the integration and semantic homogeneity between them. In order to provide the reliable knowledge (in terms of new treatments, new clinical findings, new exercises), medical ontologies should be evolved. The adaptation of ontology according to the changes in domain of interest is called ontology evolution. This paper investigates ontology evolution approaches applied in medical domain. It also presents a framework proposal for comparative study that demonstrates the difference between the approaches.

Keywords— Ontology engineering, Ontology evolution, Ontology enrichment, Ontology population.

I. INTRODUCTION

Ontology evolution is defined by Haase and Stojanovic [1] as the process that “adapt and change the ontology in a consistent way”. Plessers [2] 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”. In general the process of adaptation of the ontology according to the changes in the domain of interest is called ontology evolution.

There is a general ontology evolution cycle proposed by Fouad Zabli et al [3]. As figure 1 demonstrates, the ontology evolution tasks are: detecting the need for evolution, suggesting changes, validating changes, assessing impact and managing changes.

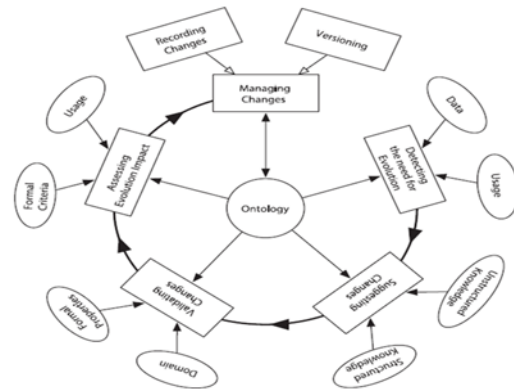


Fig. 1 Ontology evolution cycle [3]

- **Detecting the Need for Evolution:**

The need can be gotten from user conduct on a system that deals with ontology. Moreover, the need for change can be obtained from the data sources related to the ontology.

- **Suggested Changes**

The candidate changes for the ontology can be detected by either the use of unstructured knowledge (such as text documents, text corpus) or structured knowledge (such as databases and ontologies).

- **Validating Changes**

In this task, the changes that can make inconsistency in the ontology are removed. Validating changes rely on two sub tasks. The first is the domain based validation which depends on domain data sources. The second is the formal properties based validation that depends on formal techniques such as description logic that helps to detect the suggested changes which can violate the constraints defined in ontology.

- **Assessing Impact**

The goal of this task is to measure the effect of ontology changes on the external artifacts. The task consists of two subtasks:

- The effect of the changes on the operations of entities that rely on ontology
- The use of formal criteria to measure the effect of the changes. It can be applied by using formal properties.

- **Managing Changes**

The goal of this task is to save the changes that are occurred on ontology and manage different versions of ontology. Managing changes involve two subtasks:

- Recording changes that can be proceeded in a manual or

Mariam Gawich is with the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt (e-mail: mariamjawich@gmail.com).

Marco Alfonse is with the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt (e-mail: marco@fcis.asu.edu.eg).

Mostafa Aref is with the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt (e-mail: mostafa.aref@fcis.asu.edu.eg).

Abdel-Badeeh M. Salem is with the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt (e-mail: abmsalem@cis.asu.edu.eg)

automatic way.

- Versioning which can be implemented by the use of change detection tools.

The objective of this paper is the investigation of ontology evolution approaches applied in medical domain. This paper is organized as follows; section 2 presents the ontology evolution approaches applied in medical domain, section 3 provides a framework proposal for comparative study and section 4 contains the conclusion and future work.

II. ONTOLOGY EVOLUTION APPROACHES IN MEDICAL DOMAIN

There are five approaches for medical ontology evolution; SemCado, DyKOSMap, Sub Ontology Evolution in a Distributed Health Care Enterprise, Mining Relation Reversals and Semi-Automated Ontology Management (SEAM) System.

A. SemCado

M.Ben Messaoud proposed a serendipitous strategy [4] for causal discovery algorithms that detect the mapping between causal Bayesian network and ontology. The algorithm assumes the use of a single domain ontology. It represents the extension of MyCado algorithm [5, 6]. It links each node provided by the Causal Bayesian Network (CBN) with each equivalent concept provided by the ontology. The objective of SemCado algorithm is to discover and reuse the knowledge located in CBN to evolve the ontology.

SemCado takes an observational data set and corresponding domain ontology as input. The algorithm maps the concept (c_i) of the ontology domain with each node (x_i) in CBN. As figure 2 demonstrates, SemCado algorithm is executed through three stages: structure learning, casual discovery process and ontology evolution.

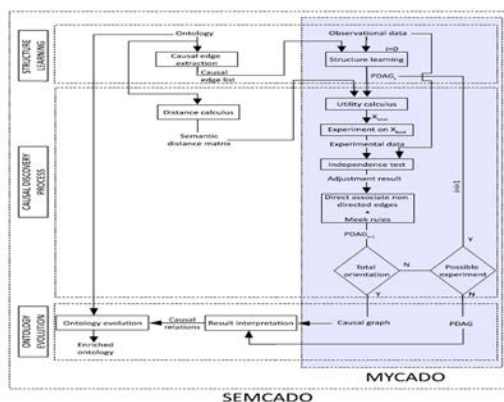


Fig.2 SemCado; Extending MyCaDo to allow CBN-ontology interactions [4]

• Learning Initial Structure Using Causal Prior Knowledge

Ontology consists of several relations between concepts such as the hierarchal and semantic relations. In order to eliminate the complexity of learning structure, the causal relations will be considered as edge constraints that will be incorporated in the structure learning process.

• Semantic Based Causal Discovery Process

Although MyCado algorithm applies the use of utility function that relies on the node connectivity which helps to detect the maximum number of edges, the edges are not considered as the most important ones. Therefore, SemCado strategy tackles this problem by applying the semantic distance calculus [7] that relies on the ontology structure.

• Edge Orientation and Ontology Evolution

The edge orientation strategy of MyCado will be applied when the Partially Directed Acyclic Graph (PDAG) contains non directed edge, a reiteration process will be executed until causal discoveries are detected. Only the causal relations, that will not violate the semantic consistency, will be obtained and mapped into semantic causal relations between the similar concepts. The output of this approach is a CBN with suggested new causal relations and enriched ontology.

The major advantage of this approach is the reusability of causal relations of Bayesian network to support the evolution of domain ontology. Moreover, Ontology evolution will be carried out without causing axioms violation. Whereas, the disadvantages of this approach are that several constraints should be taken into consideration that are proposed during the accomplishment of ontology evolution process such as the use of single domain ontology.

B. DyKOSMap Adaptation Framework for Mapping Adaptation

DyKOSMap Framework was proposed by Julio Cesar [8]. The objective of this framework is to update the mapping between two Knowledge Organizations Systems (KOS) concepts in a way that corresponds to the evolution of these KOS. In other words, DyKOSMap indicates the correlations between KOSs evolution and the mapping evolution.

KOS consists of several conceptual models such as terminologies, taxonomies, classifications, semantic networks and ontologies. All these conceptual models are used in biomedical domain such as the MedDRA terminologies [9], NCI thesaurus [10], SNOMED [11], Foundational Model of Anatomy [12] and gene ontology [13]. The framework assumes that the evolution of KOS will immediately affect the mappings that have been executed between the adjusted KOS and another. The types of changes that can be occurred in KOS are classified into atomic and complex changes. Atomic changes involve the additions, removals of KOS concepts and attributes while complex changes involve the merge, split or combination of these changes. Knowing the types of complex

changes that can be occurred in KOS will facilitate the mapping maintenance. For this reason, an approach proposed by Hartung [14] is applied to take advantage of the identification of general descriptions of complex changes via change patterns that are determined according to ICD9-CM [15].

DyKOSMap consists of several modules; a module for the Identification of effective complex changes (CI), a Change Pattern (CP) module and Mapping Evolution (ME) module. Figure 3 illustrates the DyKOSMap framework. It takes each new published version of KOS as an input. The CI module determines the changes between KOS versions through the use of specific CPs that identifies the kind of change whether it is a merge or split case. CI module produces instances of change patterns. The ME module takes the instances of patterns and current mappings as input. ME Module relies on the predefined and heuristics to produce up to date mapping which is coherent to the evolved KOS model as well as the mapping and KOS history.

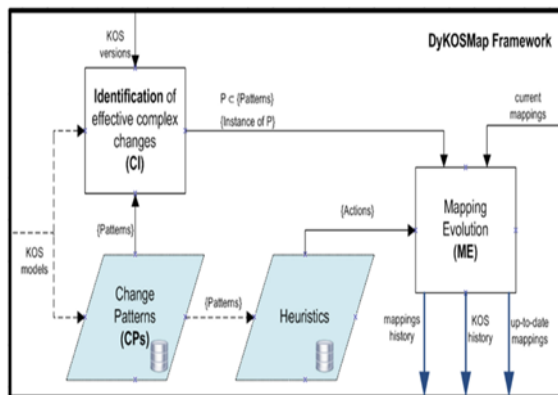


Fig.3 DyKOSMap Framework [8]

C. Sub Ontology Evolution in a Distributed Health Care Enterprise

This approach, which is referred as change propagation, proposed by Anny Kartika [16]. The approach includes a group of rules to ensure the efficiency of the generated sub ontology. Sub ontology refers to a specific part of a base ontology which is extracted to a specific context, specialty or application [17]. The objective of this approach is to keep up the coherence between the sub ontologies with the whole ontology. The change propagation approach depends on the use of change logs of health ontologies. The approach is executed through the following phases:

• Identifying The Semantic Change Operations

Semantic change operations are determined from the change logs provided as well as the comparison between the ontologies versions. Basic semantic change operations are classified into several lists such as list of concept additions, list of concept deletions, list of relationships additions, list of deletion descriptions and list of additions descriptions.

• Change Propagation Process

Figure 4 demonstrates the propagation process. The list of semantic based change operations guides the evolution of sub ontology according to the changes occurred on the base ontology.

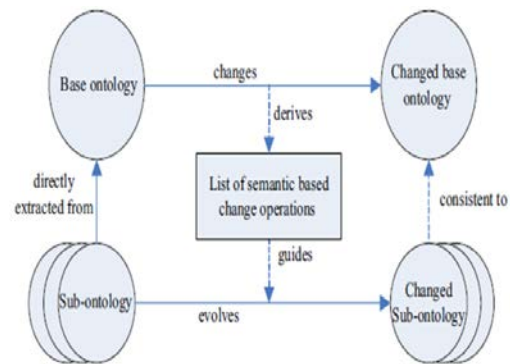


Fig.4 The propagation of change from the base ontology to the sub-ontology using the list of semantic change operations [16]

The output will be the changed sub ontology which is consistent with the changed base ontology. The consistence between the changed base ontology and the changed sub ontology is realized by the use of some rules that deal with the concept and its attributes which control the selection of sub ontologies that will be influenced by the changes of base ontology.

D. Mining Relation Reversals in the Evolution of SNOMED CT using MapReduce

Siqiang Tao and colleagues proposed this approach [18] to discover the hierarchical reversal relations among different SNOMED CT versions that start from 2009 till 2014. The approach is applied by the use of MapReduce algorithm [19].

Figure 5 illustrates an example that clarifies the notion of reversal relations. There are two versions of SNOMED; the first version (07/2013) mentions the concept 'joint structure of shoulder girdle' 'is-a' 'Joint structure of shoulder region', the second version (03/2014) mentions the concept 'Joint structure of shoulder region' is a "Joint structure of shoulder girdle". The result of merging between the two versions yields to the existence of reversal relations between the two concepts.

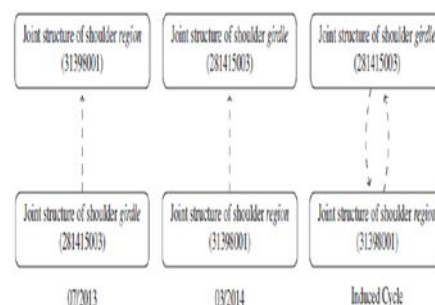


Fig.5 First two arrows: reversal of is-a relation between two versions of SNOMED CT. Right most: a cycle induced by the reversal pair when it appears in a merged graph. The numbers below concept labels are the corresponding SNOMED CT identifiers [18]

The MapReduce algorithm, which is illustrated in figure 6, is applied to extract the relation reversals hierarchy between different versions of SNOMED ontology. MapReduce algorithm involves a mapper and a reducer functions that are applied to detect the transitive closures and reversal relations. As left side of figure 6 demonstrates, the MapReduce calculates the transition closure for every SNOMED CT starting from 2009 to 2014. Transitive closure is used to discover the direct and indirect paths of the hierarchy taking into account the is-a relationship.

MapReduce takes the concept nodes and 'is a' relations as input. MapReduce introduces the hash function that uses the distributed cache to fill each concept and its parent. From lines 3 to 16 in the left side of figure 6, the map function gets the direct parent of concept C and its ancestor. From lines 17-21 in the left side of figure 6, the reducer function is applied to emit the concept and its ancestor pairs. The output of this method is a set of transitive closure pairs. In the right side of figure 6, the reversal relations are calculated using the MapReduce. Its input involves the transitive closure concept pairs that are provided by the two SNOMED CT versions O and N. The reversal relations are detected through the intersection of concept pairs and reversed concept pair provided by another version.



Fig.6 Left - MapReduce steps to compute transitive closure. Right - MapReduce steps to compute reversals [18]

The extraction of reversal relations between different SNOMED CT versions helps to visualize the changes occurred during the ontology evolution. This visualization will facilitate the integration between the different SNOMED CT ontologies. Although MapReduce focuses on the concepts which are associated with is-a relationships, it obsoletes other types of relationships.

E. Semi-Automated Ontology Management (SEAM) System

SEAM system was proposed by Kristina Harris and colleagues [20]. Its objective is to promote the extraction of information concerning a specific disease. This information can serve the ontology construction. The system takes the clinical and biomedical texts as input and generates the information in a form of clinical terms, synonyms and relationship hierarchy. Figure 7 illustrates the phases of

SEAM which are: preprocessing, term extraction, synonym and relationships extraction and candidate filtering.

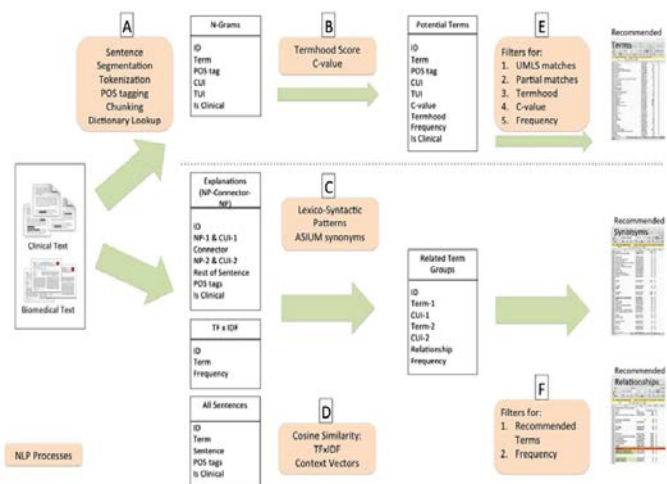


Fig.7 A Pictorial Representation of the SEAM System [20]

• Preprocessing

This phase involves the tokenization, chunking, dictionary lookup, explanation and term frequency components. The output of tokenization and chunking is a set of terms. Since the term may consist of one or more words, they are called Ngram (a term consists of two words called 2-gram). An Apache Lucene index [21] is constructed by choosing the terms from Unified Medical Language System (UMLS) metathesaurus [22]. The lookup is used to detect the concept unique identifier and semantic type identifier of the Ngram terms provided by UMLS metathesaurus. The matching between the UMLS and Ngrams terms involves full matching and partial matching. Explanation involves the extraction of patterns from sentences such as noun phrases and connectors. Finally, the Term Frequency and Inverse Document Frequency (TF-IDF) [23] are calculated to measure the similarity between the Ngram terms.

• Term Extraction

The pertinent terms that are applicable for use case are recognized. The C-value [24] and Termhood [25] algorithms are applied to discover the pertinent terms that are unnoticed.

• Synonym and Relationships Extraction

There are two pairs of synonym extraction approaches that are applied to present the concepts which are related to the use case. The first pair of synonym extraction focuses on the identification of associated topics over the corpora. Second pair of synonym extraction technique establishes lexico syntactic pattern matching [26] over the explanations. Relationships are extracted by the use of lexico syntactic that are purposed to detect the hierarchical relationships between terms.

• Candidate Filtering

SEAM filters the recommended terms, synonyms and relationships in order to facilitate the evaluation of the

candidate terms that will be established by the reviewers. Two pairs of term filters are applied; the first pair classifies the full and partial matching with UMLS metathesaurus, the second pair detects the unmatched terms by the calculation of Termhood and C-value. The terms which have the identical concept unique identifier of UMLS will be collected in a group of synonyms. For the relationship filter, the frequency calculation is applied. The frequency presents the number of times the relationship occurred.

The use of medical information sources such as UMLS as well as the use of natural language processing techniques such as tokenization, chunking, explanation and lexico syntactic patterns facilitate and enhance the recommendation of candidate terms. The lexico syntactic patterns are used to identify the relationships between the extracted terms. Moreover, the approach didn't present how the valid term will be added in the ontology. The ontology evolution depends on the expert entry.

III. FRAMEWORK PROPOSAL FOR COMPARATIVE STUDY

The points that are used to point out differences between ontology evolution approaches belongs to the following ontology evolution tasks:

- Consistency verification [27]

It involves the use of rules or constraints to eliminate the changes that can cause ontology inconsistency.

- Suggesting change

The changes for the ontology can be provided by unstructured or structured knowledge.

- Change implementation [28]

It enables the ontology engineer to accept or reject the suggested changes. Moreover, ontology engineer should be notified about the results of a change request on the ontology and its dependent application.

Table 1 demonstrates the comparison between ontology evolution approaches.

Table 1. Comparison between the Ontology Evolution Approaches

Approach Name	Ontology Evolution tasks			Kind of changes
	Consistency Verification	Suggesting for Change	Change implementation	
SemCado Approach	Yes	Depends on Causal Bayesian Network	Yes	Basic
DyKOSMAP Approach	Yes	Depends on the changes occurred on KOS models	No	Complex
Sub Ontology Evolution in a distributed health care	Yes	Depends on change logs and comparison between ontologies	Yes	Basic and Complex
Mining Relation Reversals Approach	Yes	No	No	Doesn't mention
SEAM	Yes	Depends on clinical and biomedical text as well as UMLS	No	Doesn't mention

For the SemCado approach; its consistency verification relies on the use of semantic distance calculus to select the most informative edges and the intervention of consistency rules to eliminate the inconsistent causal conclusions from ontology axioms. The evolution of the ontology relies on the causal Bayesian network which is an external structured data sources. Concerning the change implementation, SemCado displays all the consequences that can be occurred on the ontology. It enables the ontology engineer to accept or reject the changes. SemCado focuses on the ontology enrichment that involves basic changes such as the addition and modification of concepts, properties and relations

For DyKOSMap, its consistency verification depends on the use of predefined heuristic rules to generate an evolved mapping coherent to evolved KOS model. Since the CI module detects the changes that are occurred on KOS versions, the suggesting for change depends on structured knowledge. Although the output of DyKOSMap is the evolved mapping between KOS which can help the users who maintain the mapping, the approach doesn't mention the application of change implementation task. Since DyKOSMap deals with the merge and split change patterns, it deals with complex changes.

For Sub ontology evolution approach, consistency verification is applied by the use of set of rules to select the sub ontology that will be influenced by the change in the whole ontology. The suggesting of changes relies on structured data provided by log files as well as the comparison between ontologies. Concerning the change implementation, Sub ontology evolution approach connects the sub ontology concepts to archetype terms. It suggests archetype terms to the archetype author who can accept or reject them. Sub ontology deals with basic and complex changes.

Concerning the mining reversals relations, it ensures the consistency verification by its usage for MapReduce algorithm to detect reversal relations between two SNOMED versions. Mining reversals relations approach doesn't suggest changes. Although map reducer helps the users to visualize the changes between SNOMED versions, it doesn't execute the change implementation task. Mining reversal relations approach detect the reversal relations between two ontology versions but it doesn't represent them in an ontology as complex changes

Concerning SEAM approach, consistency verification is executed by the term filter, term frequency and lexico syntactic pattern to identify relationships as well as the use of UMLS to provide the pertinent terms. Suggesting changes relies on the use of unstructured data (clinical and biomedical text) and structured data (UMLS). SEAM doesn't execute the change implementation task. SEAM determines the changes in the domain but it doesn't represent them as basic or complex changes in ontology.

IV. CONCLUSION AND FUTURE WORK

The notion of ontology evolution in SemCado and Sub Ontology Evolution approaches focus on the creation of evolved ontology. Other approaches focus on the detection of inconsistency that can be occurred in the changes between two ontology files. The existence of conflicts can violate the ontology consistency and its detection will facilitate the integration between ontologies. Only SEAM approach relies on structured and unstructured data resources which provide recent changes in the medical domain. One of the interesting issues remaining for future work is to propose a model for ontology evolution that introduces the use of relevant medical data sources and takes into account the ontology consistency.

REFERENCES

- [1] P.Haase, L.Stojanovic, Consistent evolution of OWL ontologies, In Proceedings of the Second European Semantic Web Conference, Heraklion, Greece, PP. 182–197,(2005)
- [2] P.Plessers, O.De Troyer, S.Casteleyn, Understanding Ontology Evolution: A Change Detection Approach. Web Semantics: Science, Services and Agents on the WWW, 5,1, PP 39–49 (2007)
- [3] F.Zablith, G. Antoniou, M.d'Aquin, G. Flouris, H.Kondylakis,E.Motta,D.Plexousakis,M.Sabou, Ontology evolution: a process-centric survey. The Knowledge Engineering Review, 30,1, PP.45-75 (2015)
- [4] M. B. Messaoud, L. Philippe, B. A. Nahla, SemCado: a serendipitous strategy for causal discovery and ontology evolution, Knowledge-Based Systems, 76, PP. 79–95 (2015)
- [5] S. Meganck, Towards an Integral Approach for Modelling Causality, Ph.D. Thesis, INSA Rouen and Vrije Universiteit Brussels (2008).
- [6] S. Meganck, P. Leray, B. Manderick, Learning causal bayesian networks from observations and experiments: A decision theoretic approach. In International Conference on Modeling Decisions for Artificial Intelligence , PP.58-69, Springer Berlin Heidelberg (2006)
- [7] E. Blanchard, M. Harzallah, H. Briand, P. Kuntz, A typology of ontology-based semantic measures., in: 2nd INTEROP-EMOI Open Workshop on Enterprise Models and Ontologies for Interoperability at the 17th Conference on Advanced Information Systems Engineering (CAISE'05), 160, PP.407–412, 2005.
- [8] J.C.Dos Reis, C.Pruski, M.Da Silveira, C.Reynaud-Delaître, DyKOSMap: A framework for mapping adaptation between biomedical knowledge organization systems. Journal of biomedical informatics, 55, PP.153-173.
- [9] <http://www.meddra.org/> Access date, February 2016.
- [10] <https://ncit.nci.nih.gov/ncitbrowser/>. Access date, February 2016.
- [11] https://www.nlm.nih.gov/research/umls/Snomed/snomed_main.html. Access date, February 2016.
- [12] <http://sig.biostr.washington.edu/projects/fm/>. Access date, February 2016.
- [13] <http://geneontology.org/>. Access date, February 2016.
- [14] M.Hartung, A.Groß, E. Rahm, COnto-Diff: generation of complex evolution mappings for life science ontologies. J Biomed Inform. 46(1) (2013)
- [15] <http://www.cdc.gov/nchs/icd/icd9cm.htm>. Access date, February 2016.
- [16] A.K.Sari,W. Rahayu, M.Bhatt, An approach for sub-ontology evolution in a distributed health care enterprise. Information Systems, 38(5),PP. 727-744 (2013)
- [17] D.Kang, B.Xu, J.Lu, P.Wang, Y.Li, In:OTM Confederated International Conferences" On the Move to Meaningful Internet Systems". Springer Berlin Heidelberg,PP.731-740 (2004)
- [18] S.Tao, L.Cui,W. Zhu, M.Sun, O.Bodenreider, G-Q.Zhang, Mining Relation Reversals in the Evolution of SNOMED CT Using MapReduce. AMIA Summits on Translational Science Proceedings, PP.46-50 (2015)
- [19] J.Dean, S. Ghemawat, Mapreduce: simplified data processing on large clusters. In OSDI'04: Proceedings of the 6th conference on Symposium on Operating Systems Design & Implementation (Berkeley, CA, USA, 2004), USENIX Association (2004)
- [20] K.Doing-Harris, Y.Livnat, S.Meystre, Automated concept and relationship extraction for the semi-automated ontology management (SEAM) system,Journal of biomedical semantics 6.1 (2015)
- [21] <https://lucene.apache.org/core/> Access date: March 2016.
- [22] 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/>. Access Date, March 2016.
- [23] G.Salton,M.McGill, Introduction to Modern Information Retrieval. McGraw-Hill,New York (1983)
- [24] KT Frantzi, S Ananiadou, J Tsujii, The c-value/nc-value method of automatic recognition for multi-word terms". In: International Conference on Theory and Practice of Digital Libraries. Springer Berlin Heidelberg,PP. 558–604.(1998)
- [25] Q.Zeng,T.Tse,G.Divita,A.Keselman,J.Crowell, A.Browne, S. Goryachev,L.Ngo, Term Identification Methods for Consumer Health Vocabulary Development J Med Internet Res;9(1):e4 (2007)
- [26] K. Liu, W. W. Chapman,G. Savova3, C. G. Chute3, N. Sioutos4,R. S. Crowley,Effectiveness of Lexico-syntactic Pattern Matching for Ontology Enrichment with Clinical Documents. Methods Inf Med. 2011;50(5),PP.397–407 (2011)
- [27] P.Plessers, O.De Troyer, Ontology change detection using a version log, Proceedings of the 4thInternational Semantic Web Conference (ISWC-05), PP. 578-592.(2005)
- [28] L.Stojanovic,Methods and Tools for Ontology Evolution. PhD thesis, University of Karlsruhe (2004)



Mariam Gawich is a PhD student at the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt.



Dr. Marco Alfonse is a Lecturer at the Faculty of Computer and Information Science, Ain Shams University, Cairo, Egypt. He got Ph.D. of Computer Science since August 2014, University of Ain Shams. His research interests: Semantic Web, Ontological Engineering, Medical Informatics, and Artificial Intelligence. He has published around 20 publications in refereed international journals and conferences.



Prof. Mostafa Aref is Professor of Computer Science, Ain Shams University, Cairo Egypt. He got Ph.D. of Engineering Science, June 1988, University of Toledo, Toledo, Ohio, USA. He has more than 50 journal and conference publications. His research areas are Natural Language Processing, Knowledge Representation, Object-oriented Programming, Ontology and Real-Time Strategy Games.



Prof. Abdel-Badeeh M Salem He is a Professor of Computer Science since 1989 at Ain Shams University, Egypt. His research includes intelligent computing, knowledge-based systems, biomedical informatics, and intelligent e-learning. He has published around 250 papers in refereed journals and conferences. He has been involved in more than 400 Confs and workshops as a Keynote Speaker, Scientific Program Committee, Organizer and Session Chair. He is a member of many national and international informatics associations.