



An Automated Approach for Medical Ontology Maintenance System

Norah Farooqi¹, Mariam Gawich², Mohamed Osama Khozium³

¹College of Computer and Information Systems, Umm Al-Qura University, Makkah, Saudi Arabia.

²French University in Egypt (UFE)

³ Prof., Faculty of Public Health and Health Informatics, HITM Dep., ICRS Consultant; Umm Al-Qura University, Makkah, Saudi Arabia.

nsfarooqi@uqu.edu.sa, Mariam.gawish@ufe.edu.eg, mokhozium@uqu.edu.sa

Abstract: The medical domain is characterized by the frequent new changes in terms of new treatments, new clinical findings. There is a necessity to maintain the medical ontology that is used in the healthcare enterprises to present the updated information to the patient, physician and other healthcare personnel. Challenges that face the evolution of the ontology are the consistency during the evolution process and the automation of the evolution process that doesn't rely on the intervention of the user. Furthermore, the medical ontology lacks from the use of the unified standardized medical concepts. This paper presents a new approach that executes evolution in an automated way which ensures the ontology consistency. The proposed automated ontology evolution system is applied on the prior researched medical ontology evolution system. Moreover, the scientific standardized term is used instead of the commercial or localized term.

[Norah Farooqi, Mariam Gawich, Mohamed Osama Khozium. **An Automated Approach for Medical Ontology Maintenance System.** *Life Sci J* 2020;17(11):24-31]. ISSN: 1097-8135 (Print) / ISSN: 2372-613X (Online). <http://www.lifesciencesite.com>. 4. doi:[10.7537/marslsj171120.04](https://doi.org/10.7537/marslsj171120.04).

Keywords: Knowledge engineering, medical informatics, ontology management, ontology evolution, ontology maintenance.

1- Introduction

Several researchers have defined the aspect of ontology evolution. Some researchers [1] have defined it as “the process of adaptation and modification of the ontology with a consistent manner”. Other researchers [2] have defined the ontology evolution activity as “the process of adapting the ontology according to new changes in the domain taking into consideration the consistency of the ontology and the consistency of the applications that deal with the ontology”. The goal of the ontology evolution is to update the ontology by keeping its consistency and relevance to changes in the domain of interest. Almost of researchers consider the notion of the ontology evolution is the same notion of ontology maintenance except Maedche and Volz [3] differentiate between the two notions. The ontology evolution can be a part of the ontology management system that covers almost of the ontology engineering activities such as the ontology mapping, ontology alignment, ontology annotation and ontology versioning.

There is a centric survey for the ontology evolution [4] proposed by Zabliith and other researchers. They proposed a life cycle for the ontology evolution presented in figure 1. It includes five main tasks: detecting the evolution need, changes

suggestion, changes validation, assessing impact and managing changes.

For the first task which is the detection of the need for the evolution. It can be determined by the user or the change in the data sources that necessitate the modification of the ontology. For the second task which is suggesting changes, the changes are suggested from different data sources. These sources can be structured and unstructured data.

For the third task which is the validation of changes. It can be realized by the use of accredited data sources specialized in the domain of the ontology. In addition, the use of constraints or formal properties can be used to keep the consistency of the ontology.

For the fourth task which the assessing the effect of a change on the current ontology. It means the influence of the modified ontology elements on the depending applications that are connected to the evolved ontology. For the fifth task which is the management of the changes, it can be established by the creation of the log file that registers all the modifications that are applied on the ontology. In addition, the change detection tools can be used in this task to facilitate the visualization and the management of different ontology versions.

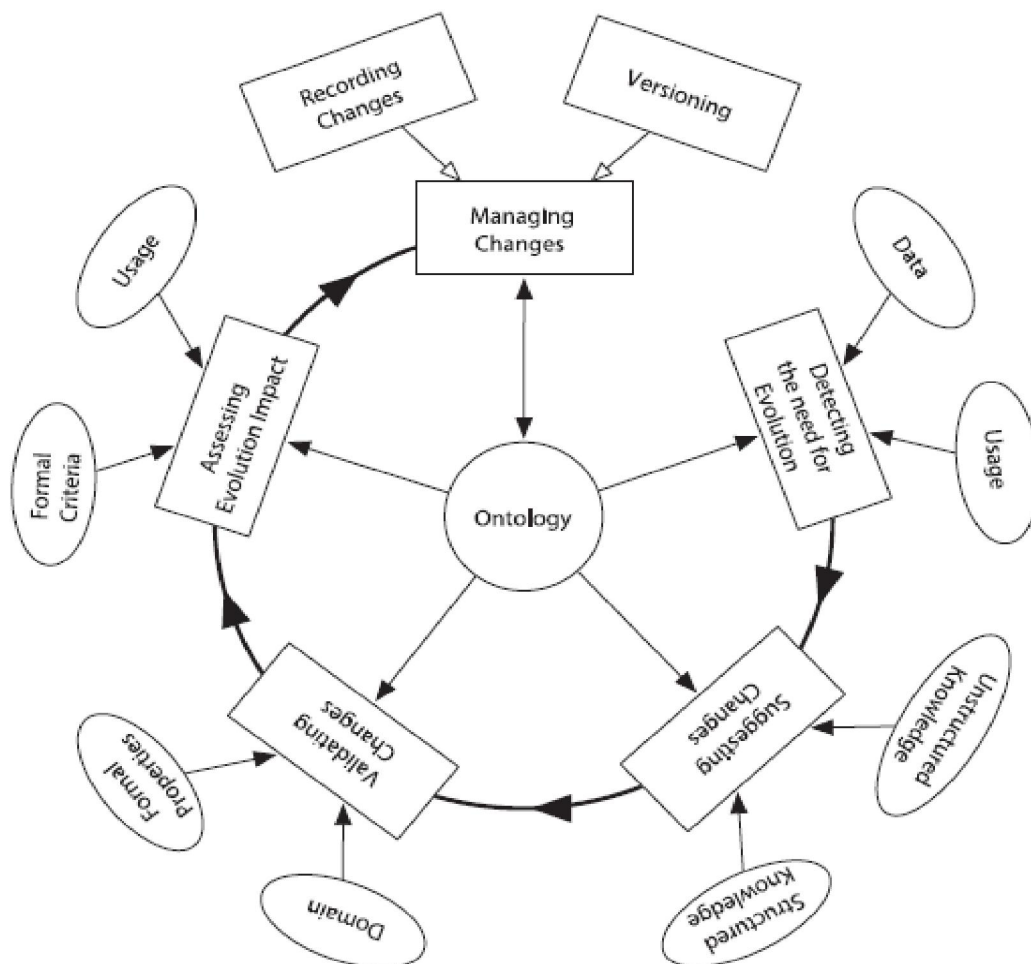


Figure 1- Ontology Evolution Life Cycle [4]

The objective of this paper is to present an extension for the ontology evolution system [5] that is used to evolve the medical ontology taking into account its consistency. The research paper include the following sections; section 2 presents the related works, section 3 presents the proposed ontology evolution system, section 4 presents the implementation of the ontology evolution system on the rheumatoid ontology, section 5 provides the evaluation of the proposed system and section 6 consists of the conclusion and future work.

2- Related work

In the medical and biomedical domains, various researchers presented the approaches to evolve the ontology. M. Ben Messaoud et al [6] proposed the SemCado approach to evolve biomedical ontology. The SemCado is the extension of the MyCado algorithm [7] that builds the Causal Bayesian Network (CBN) from biomedical observational dataset. The SemCado takes the nodes that are located in the CBN

and maps them to the ontology classes and subclasses. Furthermore, SemCado applies the use of semantic calculus [8] to detect the most informative semantic causal relations provided by the CBN. These semantic causal relations will be validated by the experts before the execution of the evolution process. The outputs of this approach is a CBN built from the observational dataset as well as the evolved ontology based on the discovered semantic causal relations.

Siqiang Tao and other researchers [9] presented an approach that mines the reversal relations between classes that belong to two ontologies versions. The objective of the approach is the identification of the transitive closure and the reversal relations that can be found between several Systematized Nomenclature of Medicine (SNOMED) [10] versions. The identification of the reversal relations depends on the use of MapReduce algorithm [11] that takes two SNOMED versions as input. The inputs of the MapReduce algorithm are the ontology concepts and the is-a object properties found in the two SNOMED

versions. The Mapreduce algorithm identifies the transitive closure pair and the reversal relations that are occurred in two SNOMED versions. The Mapreduce displays to the user the inconsistencies that are occurred in terms of reversal changes executed during the ontology evolution.

Kristina Harris and other researchers [12] presented an approach to evolve the biomedical ontology called the Semi Automated Ontology Management (SEAM). Its objective is to extract set of terms, its variants and its relations provided by a biomedical text corpus specialized in a specific disease. In order to extract the terms from the corpus, SEAM applies the tokenization, chunking, term frequency, calculation of the term hood [13] and the C-value [14]. SEAM connects the Unified Medical Language System (UMLS) met thesaurus [15] to extract the synonyms of each term provided by the corpus as well as its related terms. Concerning the extraction of the relations between the terms, SEAM applies the lexico syntactic pattern matching [16] to identify the hierarchical relations between the extracted terms. The output of the SEAM include recommended common terms, common synonyms and relations that will be used to evolve medical and biomedical ontologies.

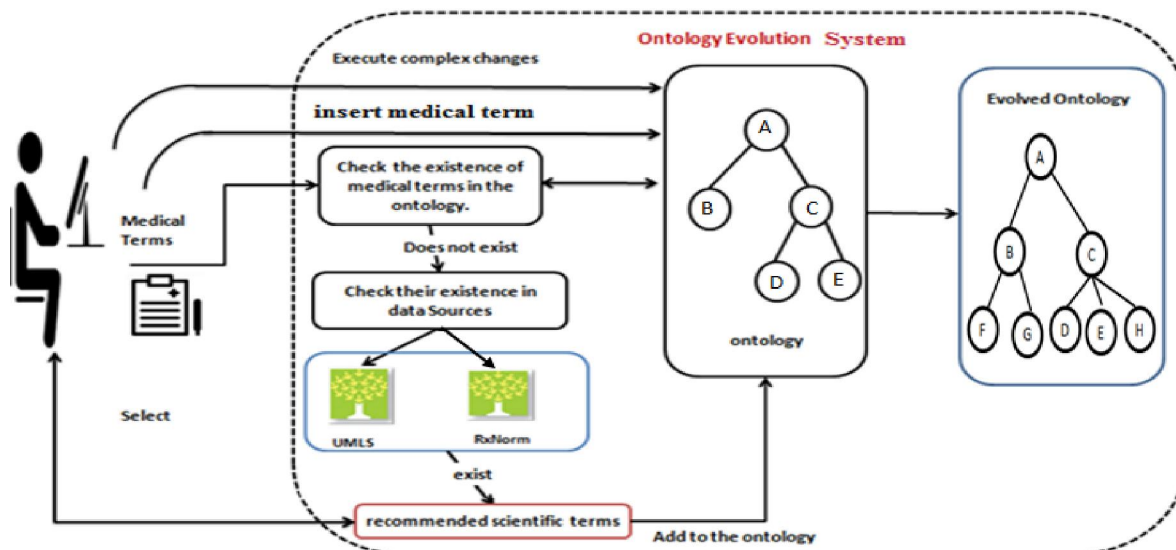
Anny Kartika and other researchers [17] presented an approach to evolve subontology that is applied in a distributed health care enterprise. The process of the evolution relies on the use of the updated log file of the whole health ontology. The inputs contain an entire health ontology and its change

log file. The approach determines the semantic operations that are executed on the whole health ontology. These semantic changes operations are identified through the comparison of the matching between the current whole health ontology and the change log file. The output of this approach is an updated subontology that is consistent with the whole health ontology.

DyKosMap [18] is a framework to evolve the mapping between two Knowledge Organizations Systems (KOS). The KOS can be a terminology, semantic networks, classifications and ontologies. The objective of the DyKosMap is to evolve the mapping between the KOS based on the evolved KOS. The framework identifies the updated mapping based on the use of Hartung change patterns [19] and some heuristic rules. The framework identifies the complex changes through the change patterns as well as the basic changes that can be determined by the heuristic rules.

3. Proposed Automated Ontology Evolution System Architecture

The proposed system architecture is the extension of the research system [5] presented in figure 2. It contains three components: the data sources, the recommendation of the scientific name and the ontology evolution process. The evolution process in the researched system is executed in a semi automatic way. The system involves the execution of the automatic evolution.



A: Rheumatoid B: Treatment C: Clinical Findings D: Signs
 E: Symptoms F: Conventional DMARDS G: biological DMARDS H: Laboratory tests

Figure 2- Ontology evolution System architecture [5]

3.1 Medical Data sources

A- UMLS met thesaurus

UMLS was created by the National Library of Medicine (NLM) [20]. It involves three components; the UMLS met thesaurus, the specialist lexicon and the semantic browser.

In UMLS met thesaurus, each concept is attached to one of the semantic types [21]. The UMLS met thesaurus is a semantic network that includes various data files [22]. The MRCONSO.RRF is one of these data files that include the medical concepts names as well as their equivalent. The MRCONSO.RRF is converted to a database file to facilitate the connection between the other system components. The objective to use the UMLS met thesaurus in the system is to recognize the medical term that can be verbalized in many ways. The medical term involves treatment, signs, symptoms, clinical finding and medical procedures.

B- Rxterms

A medical terminology interface [23] is created by the NLM. It encompasses the RxNorm [24] which refers to the standardized nomenclature for clinical drugs. The Rxterms is a specialized drugs database that involves the drug name in its normalized form, its branded drug name and its full generic name. The objective to use the Rxterms in the system is to recognize the treatment name that is not found in the UMLS met thesaurus. Furthermore, The RxNorm publishes each month an updated version of Rxterms whereas the UMLS publishes its updated version each year.

3.2 Recommendation of scientific names

This component allows to the user to insert a medical term to the ontology. The component will verify the existence the medical term in the ontology. This step is established to ensure the ontology consistency. In case that the user inserts a commercial treatment name or the Acronym of a medical concept, the component will connect the medical data sources to get the scientific standardized form for the medical term. Furthermore, the component will verify again the ontology to ensure that the ontology doesn't contain the medical term with its scientific form. Finally, the component will present only the scientific name to the user who can decide to add it in the medical ontology as an individual that will be attached to a specific class or to add it as a new class in the ontology. In case that the user inserts a term that is not found in the medical data sources, the component disables the user to insert the term to ensure the coherence of the updated elements with the medical domain.

3.3 Ontology Evolution Process

A- Semi automatic Ontology evolution process

For the semi automatic evolution, the researched system allows to the user to execute the ontology evolution. Furthermore, the system allows to the user to execute customized basic and complex changes on the current ontology. The basic changes involve the addition or the removal of ontology elements: Classes, individuals, object properties and data properties. The complex changes involve the merging, the splitting and the moving of the classes in the ontology. The system ensures that each term inserted by the user is fully related to the medical domain. Since the medical term can be expressed by different names, the system connects the data sources to retrieve the scientific standardized form of the medical term. For example: if the user inserts a treatment called Avara, the system will obligate him to use the equivalent scientific term which is leflunomide.

The evolution process involves two kinds of change operations; the basic changes and the complex changes. For the basic changes, the user can add the recommended scientific name as an individual. The evolution process component will present all ontology classes to the user who can select the class that will include the suggested scientific name as an individual. In case that the user decides to insert a new class and assert it with the suggested scientific name as an individual, the component makes sure that the new class inserted by the user doesn't exist in the ontology. Furthermore, the component enables the user to perform other basic changes such as the addition, modification and removal of classes, subclasses, individual, data properties and object properties. For the complex changes, the ontology evolution component allows to the user to perform the splitting, merging, moving operations. The component keeps the individuals, object properties and data properties that are asserted to the class on which the user decides to execute a complex change operation on it.

B- Automatic Evolution Process

As figure 3 demonstrates the proposed ontology evolution system, it takes only the current ontology as input and compares it to the ontology. The automatic ontology evolution system enables the user to apply the automatic evolution which is based on the comparison between the current ontology and the data sources. The system takes the current ontology as input and extracts all the ontology elements in terms of classes and individuals to compare them with the UMLS met thesaurus (only the concepts that are relevant to the medical ontology are considered). The comparison takes into consideration the synonyms of each class and individual through the use of the UMLS

met thesaurus and the Rxterms. If the evolution system detects new medical terms that neither they nor their synonyms exist in the current medical ontology, the

system suggests adding them in the current ontology. each medical term that is added in the ontology is presented with its scientific name.

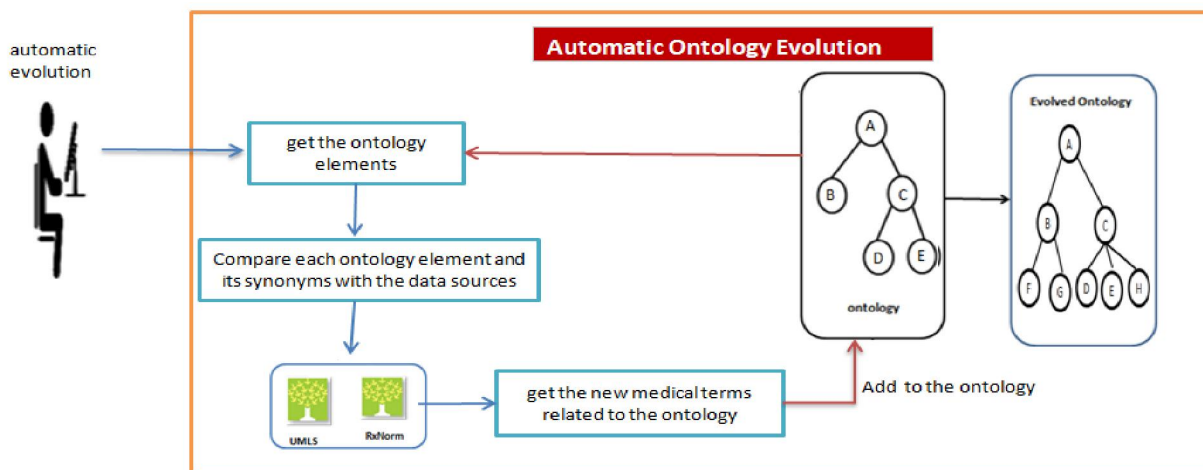


Figure 3- The Proposed Automatic Ontology Evolution System

4. Automatic Ontology evolution For the Rheumatoid Ontology

The automatic Evolution system is applied on the rheumatoid ontology [25]. The ontology evolution subsystem detects all the relevant new medical terms that are not provided by the current ontology. The Rheumatoid disease is considered as a type of

autoimmune diseases that share common treatments and clinical findings. The evolution process focuses on the detection of the medical term from the UMLS and Rxterms. These treatments are related to the autoimmune diseases and the rheumatoid disease. Table 1 demonstrates the evolution results.

Table 1: The Results Of The Evolution For The Rheumatoid Ontology

Initial Ontology	Before Evolution	After the Evolution
The Developed Rheumatoid Ontology	18 classes 95 individuals 6 object properties	19 classes (semi automatic vitamin class) 99 individuals (vitamin D “semi automatic”, anti rheumatic agents, anti rheumatic preparation and Mercaputorine) 6 object properties

The ontology evolution subsystem suggests three medical terms: Anti rheumatic agents, Antirheumatic preparation and Mercaputorine. Moreover, the evolution subsystem suggests adding the Addison disease (which is medical term that is related to the Rheumatoid disease) to the ontology. The three experts confirm that the terms “Addison disease”, “Antirheumatic preparation”, “Antirheumatic agents” are relevant to the Rheumatoid and they exclude the term Mercaputorine.

5. The Evaluation of the Automatic Ontology Evolution System

The evaluation of the system relies on its comparison with other systems. Table 2 shows this comparison that uses the following criteria:

- Input.
- The use of the change log files.
- The evolution method: it refers to the way in which the evolution is executed. It can be executed in manual way, automatic way or semi automatic way.
- The ontology evolution type [26]:

Ontology evolution has two types; the ontology enrichment and the ontology population. The ontology

enrichment is the updating of the ontology in terms of new concepts, object properties and data properties. The ontology population refers to the updating of the ontology in terms of individuals.

For the SemCado approach, the inputs involve the biomedical observational dataset and a biomedical ontology. The evolution process is established in an automatic way. The SemCado didn't depend on the change log files to guide the evolution process. The SemCado executes the ontology enrichment notion and ignores the ontology population notion. The outputs are the evolved ontology and the CBN.

Concerning the mining reversal relations, it has two couple of inputs; the first couple includes the

concept node and the is-a relation. The second couple includes the transitive closure which are the concept pairs provided by two SNOMED versions, The outputs are the reversal relations and transitive closures. The approach didn't rely on the use of change log file to determine the ontology inconsistency. The consistency verification is executed in an automated way. Since the mining reversal approach focused on the consistency of the ontology after its evolution, both of ontology enrichment and ontology population are not executed. The outputs involve reversal relations and transitive closures.

Table 2: Comparison of the Proposed System with the Other Ontology Evolution Systems According To Evolution, Use of Change Log Files and Ontology Evolution Type

Criterion/ Approach Name	Evolution	Use of change log files	Ontology evolution type	
			Enrichment	Population
SemCado	Automatic	No	Yes	No
Mining relation reversals	-	No	No	No
SEAM	Automatic	No	No	No
Sub ontology evolution	Automatic	Yes	Yes	No
DYKOSMAP	Semi automatic	No	Yes	Yes
The Research ontology evolution system [5]	Semi automatic	No	Yes	Yes
The proposed automatic ontology evolution system	Automatic	NO	Yes	Yes

The SEAM approach considered the clinical and biomedical text as inputs. SEAM executes the ontology evolution in an automated way. It didn't use the change log file to determine the evolved patterns in the ontology. Although there is no execution for the ontology enrichment and ontology population, the outputs of the approach can be used to evolve the ontology. The outputs are a set of recommended terms, synonyms and relationships.

For the subontology evolution approach, the inputs involve the whole ontology and its change log file. The evolution process is executed in an automatic way. It is relied on the use of change log files to determine the evolved patterns in the ontology. it executes only the ontology enrichment. The output is a set of updated health subontologies.

For DyKOSMap approach, the inputs involve the evolved KOS versions and the old KOS mapping. The mapping evolution is established in an automatic way. It didn't rely on the use of the change log files to determine the evolved patterns in the ontology. The evolution of the mapping focused on the enrichment

process and the population process. The output is an evolved mapping of the KOS.

For the researched ontology evolution system, its input involves the current medical ontology and the medical terms inserted by the user. The evolution is established in a semi automated way. The system enables the user to implement the change on the ontology after its detection to the new scientific term name. The system didn't rely on the use of the change log file to determine the evolved pattern in the ontology. The system implements the ontology enrichment and the ontology population. The output involves an evolved medical ontology.

Concerning the proposed automatic evolution system, its inputs is a medical ontology, the evolution is executed automatically. The system doesn't enable the user to implement the change on the ontology. it doesn't depend on the use of the log file to get the updated information. The proposed system executes the ontology enrichment and the ontology population. The output involves an updated medical ontology.

6 - Conclusion and Future Work

Almost of the current ontology maintenance or evolution approaches focus on the ontology consistency after the evolution. Both of the researched ontology evolution system and the proposed automatic evolution system takes into the consideration the consistency during the evolution process. Both of the researched ontology evolution system and the proposed automatic evolution system ensure the use of the standardized medical term. What distinguishes the proposed system is the automation of the evolution process with the minimal intervention of the user and ensures the consistency of the updated ontology. Furthermore, the automatic evolution focuses on the new medical terms that are used for a specific disease relevant to the ontology. By this way, the automatic evolution keeps the coherence between the medical ontology and the domain of interest.

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 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.
- 3 A. Maedche, and R. Volz, "The Ontology Extraction and Maintenance Framework Text-To-Onto". Proceedings of the ICDM Workshop on the Integration of Data Mining and knowledge management, pp.1-12, San Jose, CA, USA, 2001.
- 4 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. 1, pp. 45-75, 2015.
- 5 Mariam Gawich, Marco Alfonse, Mostafa M. Aref and Abdel-Badeeh M. Salem. Developing a System for Medical Ontology Evolution. *Egyptian Computer Science Journal* Vol. 41, No.2. Egypt. May 2017.
- 6 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.
- 7 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.
- 8 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.
- 9 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.
- 10 <https://www.nlm.nih.gov/healthit/snomedct/index.html> [Accessed: 25 –June-2019]
- 11 J. Dean and S. Ghemawat. "MapReduce: simplified data processing on large clusters". *Communications of the ACM*, vol.51, no.1, pp.107-113, 2008.
- 12 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.
- 13 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.
- 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 https://www.nlm.nih.gov/research/umls/knowledge_sources/metathesaurus/ [Accessed:25-June-2019].
- 16 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.
- 17 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.
- 18 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.
- 19 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.
- 20 <https://www.nlm.nih.gov/> [Accessed: 30 –June-2019]
- 21 https://www.nlm.nih.gov/research/umls/META3_current_semantic_types.html. [Accessed: 30 – June-2019]
- 22 <https://www.ncbi.nlm.nih.gov/books/NBK9685/>. [Accessed: 30 –June-2019]
- 23 https://www.nlm.nih.gov/umlslicense/rxtermA_pp/rxTerm.cfm[Accessed: 30 –June-2019]
- 24 MD Clement McDonald, BE Bray, “RxTerms - A drug interface terminology derived from RxNorm”. AMIA 2008 Symposium Proceedings, pp 227–231, 2008.
- 25 M. Gawich, M. Alfonse, M. Aref and A. M. Salem, "Development of rheumatoid ontology," *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, Cairo, pp. 421-426, 2017.
- 26 A. Khattak, K. Latif, S. Lee and Y. Lee, "Ontology Evolution: A Survey and Future Challenges", International Conference on U- and E-Service, Science and Technology UNESST 2009: U- and E-Service, Science and Technology, Part of the Communications in Computer and Information Science book series (CCIS), vol. 62. Springer, Berlin, Heidelberg, pp.68-75, 2009.

11/3/2020