

Leveraging the Impact of Ontology Evolution on Semantic Annotations

Silvio Domingos Cardoso^{1,2(✉)}, Cédric Pruski¹, Marcos Da Silveira¹, Ying-Chi Lin³, Anika Groß³, Erhard Rahm³, and Chantal Reynaud-Delaître²

¹ LIST, Luxembourg Institute of Science and Technology,
5, Avenue des Hauts-Fourneaux, 4362 Esch-sur-alzette, Luxembourg
{silvio.cardoso,cedric.pruski,marcos.dasilveira}@list.lu

² LRI, University of Paris-Sud XI, Orsay, France
chantal.reynaud@lri.fr

³ Institute of Computer Science, Universität Leipzig,
P.O. Box 100920, 04009 Leipzig, Germany
{lin,gross,rahm}@informatik.uni-leipzig.de

Abstract. This paper deals with the problem of maintenance of semantic annotations produced based on domain ontologies. Many annotated texts have been produced and made available to end-users. If not reviewed regularly, the quality of these annotations tends to decrease over time due to the evolution of the domain ontologies. The quality of these annotations is critical for tools that exploit them (e.g., search engines and decision support systems) and need to ensure an acceptable level of performance. Although the recent advances for ontology-based annotation systems to annotate new documents, the maintenance of existing annotations remains under studied. In this work we present an analysis of the impact of ontology evolution on existing annotations. To do so, we used two well-known annotators to generate more than 66 million annotations from a pre-selected set of 5000 biomedical journal articles and standard ontologies covering a period ranging from 2004 to 2016. We highlight the correlation between changes in the ontologies and changes in the annotations and we discuss the necessity to improve existing annotation formalisms in order to include elements required to support (semi-) automatic annotation maintenance mechanisms.

Keywords: Ontology evolution · Semantic annotations · Life sciences

1 Introduction

The use of ontologies, or more generally speaking Knowledge Organization Systems (KOS) [1] (which includes classification schemes, thesauri or ontologies), to annotate documents, is a current practice in order to make their semantic explicit for computers. This is for instance the case in the biomedical domain where main interests for healthcare professionals to annotate documents are twofold: (1) to transfer these documents to other institutions/people (e.g., to

accelerate the reimbursement process, to request second opinion, etc.); (2) to easily retrieve patient information. Secondary uses of these annotations are often foreseen for decision support systems, public health analysis, patient recruitment for clinical trials, etc. In the biomedical field the entities annotated include diseases, parts of the body, genes, etc. [2]. There are many structured forms to represent annotations, basically the inputs and outputs from clinical documents when it is processed by software as text processors (e.g. GATE, NCBO Annotator, MetaMap) can be expressed as annotations [2]. This is usually done by associating concept code or label of a given KOS to an element of the document (see Fig. 1). Through this link, human and computers can have an unambiguous understanding of the content of the document.

However, the dynamic nature of KOS may affect the annotations each time a new version is released. Actually, new KOS concepts can be added, obsolete ones can be removed and existing concepts may have their definition refined through the modification of their attribute values [3]. In consequence, changes in concepts can alter their semantics and therefore create a mismatch between the versions of the same concept (e.g. version 1 can be more abstract or more specific than version 2) impacting the validity of the semantic annotation. Following this observation, it is important to constantly evaluate and adapt the annotations to insure an optimal use of the annotated data. Nevertheless, the revision can hardly be done manually by virtue of the huge amount of existing annotations. Therefore, there is an urgent need for intelligent tools to support domain experts in this task.

In this paper our objectives are twofold. First, we aim at quantifying the impact of KOS evolution on the associated annotations to justify the need of automatic tools for maintaining the validity of annotations over time. This is done through systematic analyses of 66 millions of annotations obtained using biomedical journal articles and 13 successive versions of two standard medical KOS: ICD-9-CM and MeSH which will complement existing reviews that usually focus on one specific ontology [4]. Second, we discuss the capabilities of existing annotation models that deal with KOS evolution and propose new key features to cope with this problem.

The remainder of the paper is structured as follow: in Sect. 2 we review related work of the field semantic annotation evolution. Section 3 describes the experiments we have conducted to obtain the results presented in Sect. 4. Section 5 discusses the results and introduces our model to deal with annotation maintenance. Section 6 concludes the paper and outlines future work.

2 Related Work

Semantic annotation is the central notion of this work. However, many definitions can be found in the literature. According to Oren et al. [5], the term annotation can denote the process of annotating as well as the result of this process. Moreover, they distinguish three families of annotations. *Informal annotations* that are not machine-readable, (e.g. a handwritten margin annotation in a book).

Formal annotations that are machine-understandable but are not defined using ontological terms, (e.g. highlights in a html document). Last, and the kind of annotation we are referring to in this paper, *ontological annotations* that are machine-understandable and are taken from an ontology (see Fig. 1).

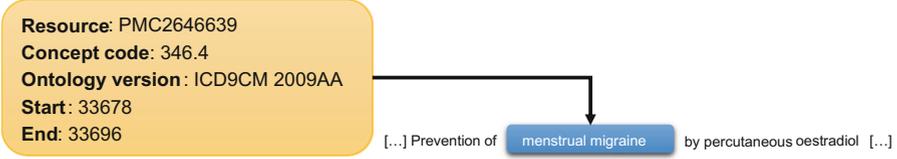


Fig. 1. Example of annotation using the concept recognition process for a PubMed document. The term *menstrual migraine* is annotated with the KOS concept 346.4 that belongs to ICD-9-CM version 2009AA (UMLS)

2.1 Existing Annotation Models

To represent annotations in the biomedical field, Luong and Dieng-Kuntz [6] defined the following annotation model:

$$\mathcal{SA} = (R_a, C_a, P_a, L, T_a) \quad (1)$$

Where:

R_a : set of resources, for instance, an RDF resource.

C_a : set of concept names defined in ontology ($C_a \subset R_a$)

P_a : set of property, for instance, an `rdf:type` ($P_a \subset R_a$)

L : set of literal values, for example, “Fever”, “Malaria Fever”, etc.

T_a : set of triples (s,p,v) where $s \in R_a$, $p \in P_a$ and $v \in (R_a \cup L)$

Gross et al. [7] and Hartung et al. [8] gave a more complete definition of an annotation, taking evolution aspect into account which was missing in Luong et al. model. In their work an annotation is defined as:

$$\mathcal{AM} = (I_u, ON_v, Q, A) \quad (2)$$

Where:

$I_u = (I, t)$: is an instance source. It consists of a set of instances $I = \{i_j, \dots, i_n\}$, e.g., molecular biological objects such as genes or proteins, at timestamp t . Instances are described by an accession ID.

ON_v : is an ontology in the version v that contains (C, R, t) , it comprises a set of concepts $C = \{c_1, \dots, c_n\}$ and relationships $R = \{r_1, \dots, r_m\}$ released at time t .

Q : is a set of quality indicators (ratings) of annotations. The quality indicators may be numerical values or come from predefined quality taxonomies, e.g., the evidence codes for provenance information or stability indicators.

A : is a set of annotations. A single annotation $a \in A$ is denoted by $a = (i, c, \{q\})$, i.e. an instance item $i \in I_u$ is annotated with an ontology concept $c \in ON_v$ and a set of quality indicators (ratings) $\{q\} \in Q$

Recently, the W3C has published a new candidate recommendation for expressing annotation¹. An annotation includes a body and a target and the relation between these two entities that may vary according to the intention of the annotation. This model is the foundation of a more general framework for sharing and reusing annotated information across different hardware and software platforms. However, this model is still not sufficient to deal with evolution issues as we will show in the following sections.

2.2 Annotation Evolution Techniques

As mentioned, the dynamic of knowledge leads to frequent revisions of KOS content which, sometimes, impacts the definition of the semantic annotations associated with documents (as illustrated in Fig. 2) [9]. The most recent approaches to analyse the evolution of the annotations is focused on biological domain, in particular on GO annotated documents. Traverso-Ribón et al. [10] developed the AnnEvol framework to compare two versions of a dataset (for instance, UnitProt-GOA and Swiss-Prot) and to verify the entities in the $dataset_{(i)}$ and $dataset_{(i+1)}$ that are similar and those which are different, using evolution criteria (e.g. obsoleted, removed and added annotations).

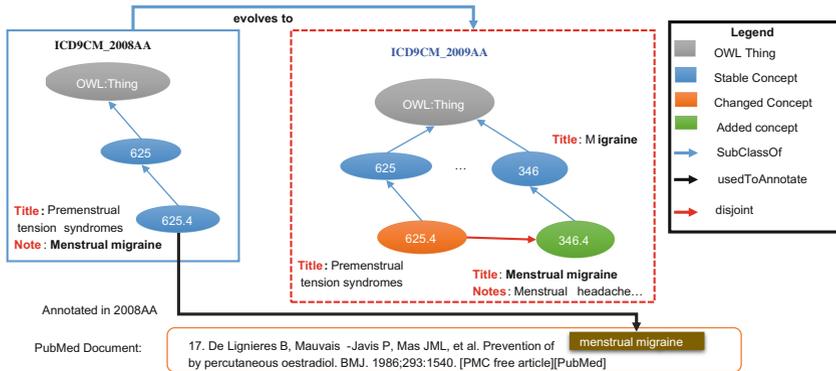


Fig. 2. Annotation evolution case study. A subset of a document is annotated with *Menstrual migraine*, an attribute of the concept 625.4 of ICD-9-CM version 2008AA. In the next version the attribute of 625.4 is removed and added as a new concept 346.4. This change has caused a mismatch between the annotation created with the older version and the concept of the new KOS version

Groß et al. [11] provide a method to test to what degree changes of GO and GO annotations (GOAs) may affect functional enrichment analyses, analyzing two real-world experimental datasets as well as 50 generated datasets.

¹ <http://www.w3.org/TR/annotation-model/>.

They proposed two types of stability measures to assess the impact of ontology and annotation changes. Differently from AnnEvol, Groß et al. deal with other change types, besides *add* and *delete*, such as, *merge* (merge of two or more categories into one category). They also verified strong structural changes as *addR* (insertion of a new relationship *r*), *delR* (deletion of an existing relationship *r*). However, these changes do not significantly impact on GOAs. As result they concluded that term-enrichment results are significantly affected by ontology and annotation evolution.

Luong and Dieng-Kuntz [6] developed the CoSWEM framework to investigated annotation evolution and explored a rule-based approach to detect and correct basic annotation inconsistencies, such as deletion. This approach converts ontologies to RDF(S) files and detects annotations affected by their evolution, as well as potentially inconsistent annotations using CORESE. Afterwards, inconsistent annotations are detected and corrected. This work focuses on expressive and small-sized ontologies and can hardly be applied to large biomedical ones, because the implemented reasoning techniques require the power of description logics (not always used in biomedical controlled terminologies) to decide on the validity of the annotations.

Frost and Moore [12] proposes a novel algorithm for optimizing gene set annotations to best match the structure of specific empirical data sources. The proposed method uses entropy minimization over variable clusters (EMVC). It filters the annotations for each gene set to remove inconsistent annotations. The results show that EMVC can filter between 92 % and 67 % of the inconsistent annotation from MSigDB C4 v4.0 cancer modules using leukemia data and MSigDB C2 v1.0 using p53 data, respectively. This method is able to improve the annotations but does not produce good results to improve incomplete gene sets or identify new gene sets. It is very sensitive to several algorithm parameters, specifically, the cluster method and it can be computationally expensive. Furthermore, the author's highlight that EMVC only works in gene set domain, thus other domains can not take advantage of this approach.

In summary, we concluded that the existing approaches to deal with annotation evolution just handle with simple changes (like concept addition and deletion), and only study the evolution of GO ontology. Furthermore, almost all of the works do not propose any method to maintain the annotations. Therefore, it is necessary to better analyze the stability of KOS annotations based on different KOS like ICD-9-CM and verify possible features to take into account to properly maintain semantic annotations in biomedical and clinical use cases.

3 Experimental Assessment of the Impact of KOS Evolution on Semantic Annotation

To bridge the gaps underlined in the previous section, we decided to conduct an empirical analysis regarding the evolution of the KOS and annotations. The lessons we learn through these experiments will allow us to come up with new proposal to deal with semantic annotation evolution issues. The used material and the adopted assessment methodology are detailed in this section.

3.1 Material

As our objective aims at analysing the evolution of semantic annotation, we have to work on several versions of an annotated corpus. Since no gold standard containing successive sets of annotated documents, we had to build our own environment. To this end, we used two annotation tools (based on distinct annotation methods), two different medical standard KOS and their associated successive versions, an ontology *Diff* tool to be able to identify the evolution of the concepts used to produce the annotations and a collection of biomedical documents. The documents were collected from the 2014 Clinical Decision Support Track (TREC 2014) campaign. It contains 733,138 biomedical articles about generic medical records. All documents from this database are open access documents from PubMed Central PMC. For our analyses we selected 5000 documents randomly.

The set of KOS is composed of several versions of medical KOS, represented in OWL format and used as “reference ontology” for text annotation. In order to annotate the documents, we selected two KOS: International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM); and Medical Subject Headings (MeSH). We collected 13 official versions of each KOS released between 2004 and 2016 in UMLS and we transformed them into OWL files.

Regarding the annotation tools, the selection criteria were: be open source, allow selecting the reference ontology, provide APIs, have good documentation, and have been extensively used for research and/or commercial purposes. We first selected General Architecture for Text Engineering (GATE) [13]. It provides support for Ontology-Aware NLP, allowing loading any ontology as RDF file and then uses a gazetteer to obtain lookup annotations that have the text offset (offset is a pair {start, end} that indicates the distance, in terms of characters, from the beginning of the document. {start} indicates the position of the first character of the text while {end} indicates position of the last character), instance and class URI. The second selected tool is the NCBO Annotator. It is part of the NCBO Annotator framework and uses a dictionary built by extracting from KOS all concepts’ label and/or other associated attributes (e.g., synonyms) that syntactically identify concepts [14]. Both annotators utilize different algorithms to produce the annotations. In this case, GATE uses Ontology-Aware NLP and NCBO Annotator uses MGrep. Moreover, NCBO Annotator also allows using other KOS to annotate the term, if a mapping exists between the concepts of both KOS. For instance, *melanoma* could also be annotated with the concept C0025202 (from NCI Thesaurus), or C0025202 (from SNOMED CT).

We used COnto-Diff [15] to determine an expressive and invertible *diff* evolution mapping between two versions of an ontology. It calculates basic change operations (insert/update/delete) from two KOS versions expressed in either OWL or OBO based on predefined set of rules defining basic and complex transformations (e.g., concept merging, concept splitting, move of concept, etc.)

3.2 Method

To identify and quantify the impact of changes affecting KOS concepts involved in annotations (as illustrated in Fig. 2), we proposed the methodology depicted in Fig. 3.

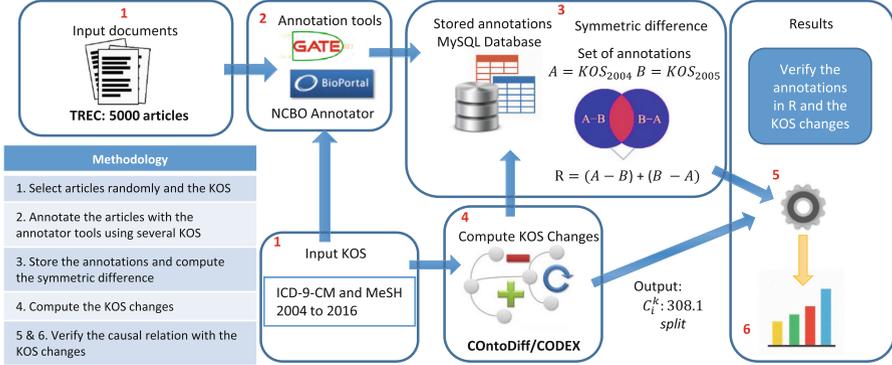


Fig. 3. The experimental protocol. The numbers in red correspond to the six steps explained in the text.

The six steps of the methodology are the following:

1. We randomly selected 5000 documents from the TREC corpus and collected the 13 KOS versions of ICD-9-CM and MeSH (from 2004 to 2016).
2. We used GATE and NCBO Annotator to annotate these documents. We configured GATE and NCBO Annotator to use one specific KOS version and repeated the annotation process for each version. We filtered the annotations produced by both annotators according to [16] (e.g., keep the longest match concept for an annotation).
3. We regrouped all annotations in one database. We then computed the symmetric difference $A_{m,n} \Delta A_{m,n+1}$ between the two annotation sets ($A_{m,n}$ and $A_{m,n+1}$) generated for a document R_m using two successive KOS versions (K_n and K_{n+1}) as the following:

$$A_{m,n} \Delta A_{m,n+1} := \{a \mid a \in A_{m,n} \wedge a \notin A_{m,n+1}\} \cup \{a \mid a \in A_{m,n+1} \wedge a \notin A_{m,n}\} \quad (3)$$

a is an annotation that can be described as $\{i, Offset, c\}$ where i is an instance at position $Offset$ annotated with a KOS concept c . The symmetric difference allows us to identify annotations that have been removed, added and modified.

4. To identify KOS changes, each pair of two KOS successive versions was input into Conto-Diff to compute the KOS difference. The difference was stored into another MySQL database and has been reused to explain the changes.

- We compared the 13 annotation sets of each document by pairs [2004–2005, 2005–2006 ...] to identify what changed in the annotations and to find correlations with the KOS changes identified by COnTo-Diff. An annotation a is considered as evolved to a' if the *Offset* or/and the c of a are different from those of a' and there is an overlap of both *Offsets*.
- Finally, we analysed the generated subset of annotations/KOS changes in order to understand the impact of KOS changes on the annotations.

4 Results

The methodology described in the previous section has allowed us to produce more than 66 millions of annotations. The amount of annotations varies according to the used annotation tools (GATE or NCBO Annotator) as depicted in Figs. 4 and 5. The difference between the two sets of annotations results from the method used to annotate the documents (they are not using only exact match). A general observation can be made based on Figs. 4 and 5.

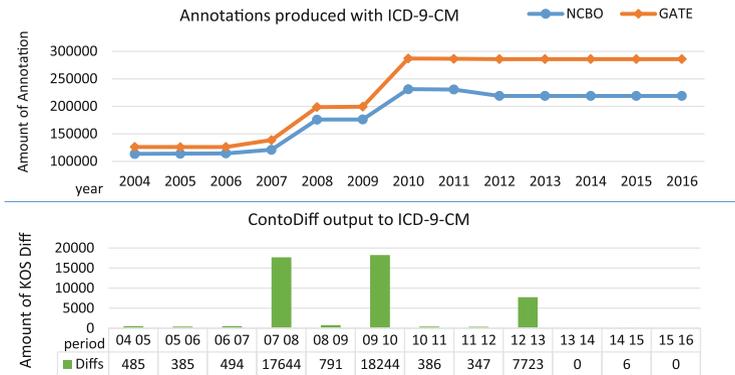


Fig. 4. Amount of annotation and KOS changes (green) produced with 13 versions of ICD-9-CM. The annotations from NCBO Annotator are represented in (blue circles) and GATE (orange diamond). The y-axis represents the amount of annotations/changes and the x-axis the KOS versions over time. (Color figure online)

We observe a huge increase in the amount of produced annotations in the periods 2007/2008 and 2009/2010 using ICD-9-CM (Fig. 4). This increase is accompanied by the changes that occurred in the KOS during these periods according to COnTo-Diff output. On the other hand, the amount of annotations in the period 2012–2013 is not increased even though there were many KOS changes. We observe an average of words/label of 8,746 during this period and thus the annotators are not able to produce annotations for these changed labels. Hence, we can conclude that the change of the number of annotations does not necessarily correspond to the amount of KOS changes. In the future work, we will

analyse what kinds of KOS changes trigger which types of annotation changes since not all kind of changes in the KOS has the same impact on the annotations (e.g., some KOS changes do not change the annotations).

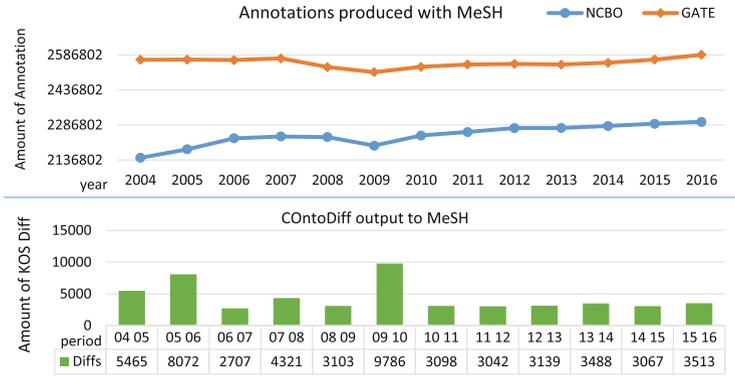


Fig. 5. Amount of annotation and KOS changes (green) produced with 13 versions of MeSH. The annotations from NCBO Annotator are represented in (blue circles) and GATE (orange diamond). The y-axis represents the amount of annotations/changes and the x-axis the KOS versions over time. (Color figure online)

In order to verify if a change in the annotations is triggered by the evolution of the KOS concepts or a gap in the annotator, we conducted the step 3 in Sect. 3.2. The first (quite evident) observation is that 100% of the annotation changes are caused by KOS changes even when the annotation methods not only produce exact matches. This simple hypothesis was not demonstrated before in the literature. We continued our analyses regarding the evolution of annotations by refining the previous sets of symmetric difference (see step 5 in Sect. 3.2). If more than one concept candidate exists to annotate a text, we used selection criteria: (1) the most recent concept and the one with largest offset, as proposed by [16]. For instance, a text with the words *chronic kidney disease* can be annotated as *kidney disease* or *chronic kidney disease*, we select only the later concept. This decision can generate changes in the annotation from one KOS version to another (change operations). One of these changes is a shift of the offsets before and after the evolution while part of these offsets overlaps. For instance, in 2007 we have the annotation “*personality disorders*”. After a KOS change in 2008 the new annotation is “*schizoid personality*” (of which “*personality*” is overlapped with the previous offset). For such case, we compute a (2) *chgOffset* operation. We formally define these conditions in Eq. (4):

$$Evolution(a_i, a_{i+1}) \longrightarrow \begin{cases} recentCp(a_i, a_{i+1}) \wedge bigOffset(a_i, a_{i+1}), & \text{if 1} \\ chgOffset(a_i, a_{i+1}), & \text{if 2} \end{cases} \quad (4)$$

As result we observe that the new KOS versions do not necessarily produce more annotations despite the increasing size of the KOS over time [9] (cf. Figs. 6

and 7). Analysing the amount of annotations and the types of changes occurring in the KOS, we observed that some minor changes which do not affect the semantics of the concepts still might impact the annotations. For instance, the concept 780.39 in ICD-9-CM version 2007AA (Seizures) evolves to (Seizure) in ICD-9-CM version 2008AA. However, both annotators did not recognize that the concepts have the same meaning and therefore the associated annotations are different from one version to the next.

We also observed that there are some periods in the KOS evolution history which are more stable and this stability is also reflected in the evolution of the annotations (e.g. the two periods 2010/2011 and 2013/2014 in ICD-9-CM on Figs. 4 and 6).

Changes in the KOS have also different impact depending on the amount of annotations a concept is associated with. This is for instance the case for the concept 084.4 of ICD-9-CM period 2007/2008 which is associated with 3143 annotations distributed in 162 documents in our corpus while concept V15.03 of ICD-9-CM period 2012/2013 is associated with only one annotation. If a single KOS change affects many annotations, it may require a huge amount of time if the maintenance of the annotation is done manually by domain experts.

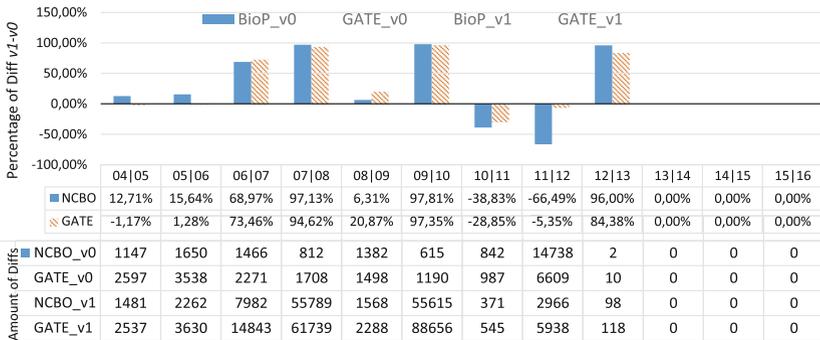


Fig. 6. Differences in two successive annotation sets produced with ICD-9-CM. The blue (solid) colour represents the annotations that belong to NCBO Annotator, and the orange (hashed) colour to GATE. (Color figure online)

We then analyse how these annotations evolve. In Table 1, we present 5 use cases showing how the annotations evolve over time and their relation with the evolution of KOS. A concept is stable if no change occurred from one KOS version to the next (second column in Table 1). In the first use case (in 2008), *hepatitis* is associated to the concept 573.3 which did not change between 2008 and 2009 (i.e. a stable concept). In 2009, another concept (571.42) was also used to annotate the term *hepatitis*. Our selection criteria define that we will select the concept with the longest title (*autoimmune hepatitis*). We also observed that this concept (571.42) changed in 2009 (a split was detected).

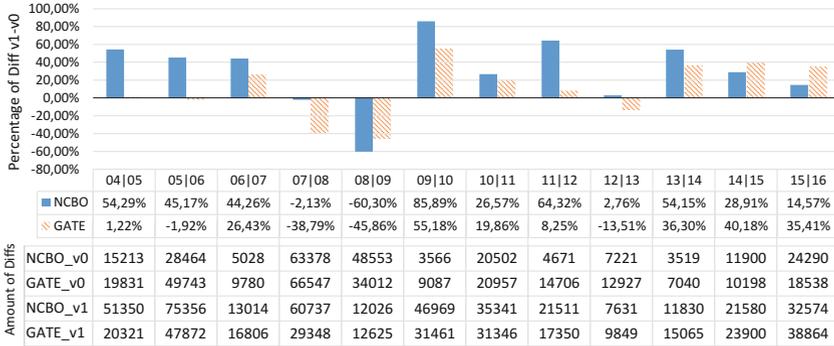


Fig. 7. Differences in two successive annotation sets produced with MeSH. The blue (solid) colour represents the annotations that belong to NCBO Annotator, and the orange (hashed) colour to GATE. (Color figure online)

The second use case illustrates a situation where both concepts changed (i.e., 625.4 had an attribute deleted, and 346.4 is a new concept).

The third use case presents the inverse situation of use case 1, i.e., an annotation evolves from a change concept to a stable concept. In a depth analysis, this case is mainly observed when more general concepts are used to annotate the text. This behaviour occurs when the annotator is not able to determine if a change in the concept has modified its meaning or not.

The last two use cases describe the addition or removal of annotations. Regarding the removal of annotations, we also verified that there are some cases where the concept remains with the same meaning, however, the annotator misses this knowledge and as result the annotation is removed from the document.

Table 1. Use cases for annotation evolution. These different cases are referred in the paper as: case 1: *stable_to_change*; case 2: *change_to_change*; case 3: *change_to_stable*; case 4: *addition*; case 5: *removal*.

Use case	KOS version	Annotation	Concept	KOS change
1	2008	Hepatitis	Change	573.3
	2009	Autoimmune hepatitis		571.42
2	2008	Menstrual migraine	Change	625.4
	2009	Menstrual migraine		346.4
3	2009	Acute renal failure	Change	584.9
	2010	Renal failure		586
4	2008	Abdominal tomography	Addition	88.02
5	2004	Bulimia	Removal	307.51

Figures 8 and 9 show how often these use cases are observed in the corpus annotated with ICD-9-CM and MeSH using GATE and NCBO Annotator, respectively. In general, we observe that changes in ICD-9-CM have less impact on the annotations than those in MeSH. The low expressiveness of ICD-9-CM can be justified as the annotators tend to apply exact match techniques for these kinds of KOS. Semantic-based techniques are more used for KOS with high expressiveness. These differences are better observed by comparing Figs. 8 and 9 to see how the annotation technique influences the final annotation results regarding to the expressiveness of the KOS. The use case 2 and 5 (*change_to_change* and *removal*, respectively) are more frequent in the MeSH based annotations. Thus, annotations based on ICD-9-CM evolve quite similarly for GATE and NCBO Annotator, while the annotations based on MeSH evolve differently, depending on the used annotator.

Taking into account the annotators techniques only, we observe that GATE also tends to preserve existing annotations while the rates of new annotations over deleted ones are quite similar for both annotators. More precisely, the rates of use cases 1 and 2 over the deleted ones (GATE has more than double of NCBO) explain the results presented in Fig. 4 (number of annotations increases faster for GATE).

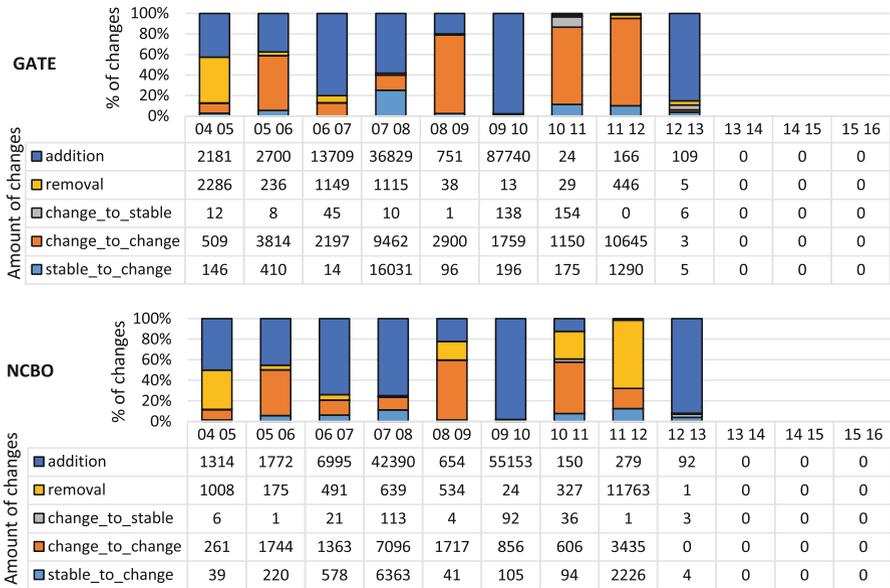


Fig. 8. Distribution of changes of ICD-9-CM annotations. The y-axis represents the percentage of changes, the x-axis the KOS versions, and below the amount of observed changes for each period is described. The listed cases follows the Table 1

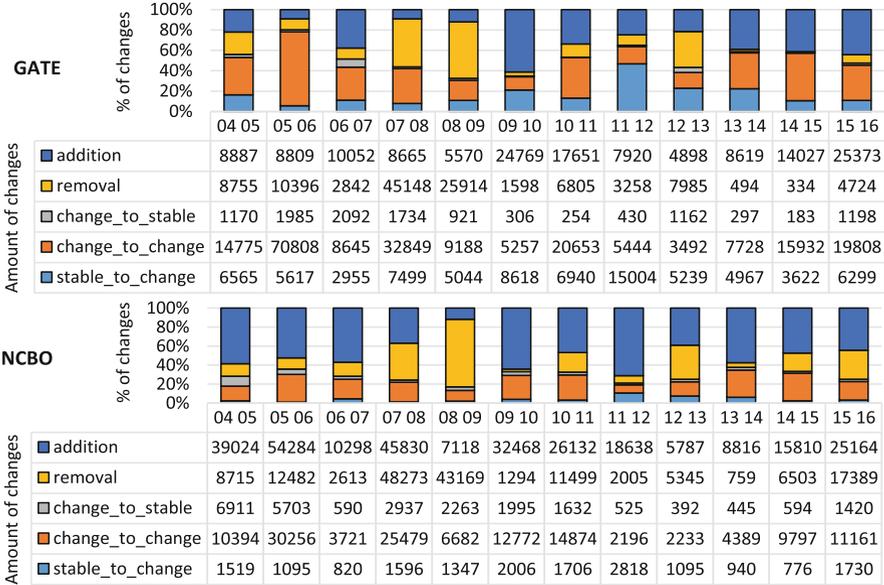


Fig. 9. Distribution of changes of MeSH annotations. The y-axis represents the percentage of changes, the x-axis the KOS versions, and below the amount of observed changes for each period is described. The listed cases follows the Table 1

5 A Model Supporting Annotation Evolution

The results presented in the previous section allow us to state that the evolution of the KOS has a direct impact on the definition of semantic annotations. However, we also showed that the modification of KOS concepts has different impacts depending on the technique that is implemented to generate the annotations. Furthermore, the evolution of KOS does not necessarily produce more information (see Figs. 6 and 7). Actually, we have observed that KOS are becoming more and more precise over time, which means the addition of new specific concepts whose labels are usually long (in terms of words) and therefore are contained very rarely in medical documents. Our study pointed out important features to take into account, at semantic annotation model level, to facilitate the maintenance of annotation over time. These features can be used to extend the model proposed by Gross et al. [7] (see Sect. 2.1). In consequence, we define our model as:

$$SAM = (I_u, ON_v, R_a, Offset, Q, H, A, SemRel, U_f)$$

Where:

- *Offset* is an element to describe the location of the element to be annotated in a given resource. From an evolution perspective, this is important for

linking annotations of different versions and also for distinguishing annotations related to the same element but are annotated differently.

- H is an element to describe which attribute of the concept (e.g., title, synonym, preferred terms, etc.) was used to produce the annotation. This element is really important since the annotation is usually defined based on the value of one concept attribute. If the corresponding concept has one of its attribute changed but not the one used to annotate, it is maybe not needed to modify the annotation.
- $SemRel$ is an element to describe the semantic relationship between the KOS concept and the annotated part of the resource. For instance, one sentence can be annotated as equivalent to a concept, more/less specific, partial match, etc. Thus, in the case of removal of a concept, the annotated sentence can be linked to the super-class of the concept and have the relation changed to “less specific”.
- U_f is an element to point to the previous version of the annotation. This element is used to keep an evolution chain of annotations.

Our proposal, allowing to link annotation versions, can also be used to improve the W3C proposal by creating an additional property called “evolved to” that links the element “annotation” to itself allowing then to create a chain of annotation version.

6 Conclusion

In this paper we made an empirical analysis of the evolution of biomedical annotations and its relation to the KOS changes. We used for that a set of documents annotated with GATE and NCBO Annotator using 13 different versions of two well-known biomedical KOS (ICD-9-CM and MeSH). We observed that there is a correlation between KOS and annotation changes. Then we regrouped the annotation changes according to the type of information that was modified and the way it was done. We obtained five different cases of changes (see Sect. 4) and verified how the annotations evolve during the KOS evolution. In a second step we analysed different annotation models in order to verify if they can represent (or if we can infer from their elements) all criteria required to classify the annotation changes. As a result of this step, we propose an extended annotation model designed to support evaluations and maintenance of annotations. However, we are still working on the maintenance methods that will use this model and other external information (e.g., KOS changes, background knowledge, etc.) to select the most adapted maintenance strategy for the annotations. We plan to continue our empirical analysis to refine the types of changes in the annotations and to determine fine grained correlations between types of changes in the KOS and types of changes in the annotations.

Acknowledgment. This work is supported by the National Research Fund (FNR) of Luxembourg and Deutsche Forschungsgemeinschaft (DFG) under the ELISA research project.

References

1. Hodge, G.: Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files. ERIC, Washington (2000)
2. Comeau, D., Doan, R., Ciccarese, P., Cohen, K., Krallinger, M., Leitner, F., Lu, Z., Peng, Y., Rinaldi, F., Torii, M., Valencia, A., Verspoor, K., Wieggers, T., Wu, C., Wilbur, W.: BioC: a minimalist approach to interoperability for biomedical text processing. *Database: J. Biol. Databases Curation* **2013**, bat064 (2013)
3. Dos Reis, J.C., Pruski, C., Da Silveira, M., Reynaud-Delaître, C.: Understanding semantic mapping evolution by observing changes in biomedical ontologies. *J. Biomed. Inf.* **47**, 71–82 (2014)
4. Hartung, M., Kirsten, T., Gross, A., Rahm, E.: OnEX: exploring changes in life science ontologies. *BMC Bioinform.* **10**(1), 1 (2009)
5. Oren, E., Möller, K., Scerri, S., Handschuh, S., Sintek, M.: What are semantic annotations? *Relatório técnico. DERI Galway* **9**, 62 (2006)
6. Luong, P.H., Dieng-Kuntz, R.: A rule-based approach for semantic annotation evolution in the coswem system. In: Kon, M., Lemire, D. (eds.) *Canadian Semantic Web. Semantic Web and Beyond*, vol. 2, pp. 103–120. Springer US, New York (2006)
7. Gross, A., Hartung, M., Kirsten, T., Rahm, E.: Estimating the quality of ontology-based annotations by considering evolutionary changes. In: Paton, N.W., Missier, P., Hedeler, C. (eds.) *DILS 2009. LNCS*, vol. 5647, pp. 71–87. Springer, Heidelberg (2009)
8. Hartung, M., Kirsten, T., Rahm, E.: Analyzing the evolution of life science ontologies and mappings. In: Bairoch, A., Cohen-Boulakia, S., Froidevaux, C. (eds.) *DILS 2008. LNCS (LNBI)*, vol. 5109, pp. 11–27. Springer, Heidelberg (2008)
9. Da Silveira, M., Dos Reis, J.C., Pruski, C.: Management of dynamic biomedical terminologies: current status and future challenges. *Yearb. Med. Inf.* **10**(1), 125–133 (2015)
10. Traverso-Ribón, I., Vidal, M.-E., Palma, G.: AnnEvol: an evolutionary framework to description ontology-based annotations. In: Ashish, N., Ambite, J.-L. (eds.) *DILS 2015. LNCS*, vol. 9162, pp. 87–103. Springer, Heidelberg (2015)
11. Groß, A., Hartung, M., Prfer, K., Kelso, J., Rahm, E.: Impact of ontology evolution on functional analyses. *Bioinformatics* **28**(20), 2671–2677 (2012)
12. Frost, H.R., Moore, J.H.: Optimization of gene set annotations via entropy minimization over variable clusters (EMVC). *Bioinformatics (Oxf. Engl.)* **30**(12), 1698–1706 (2014)
13. Cunningham, H.: GATE, a general architecture for text engineering. *Comput. Humanit.* **36**(2), 223–254 (2002)
14. Whetzel, P.L., Noy, N.F., Shah, N.H., Alexander, P.R., Nyulas, C., Tudorache, T., Musen, M.A.: BioPortal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications. *Nucleic Acids Res.* **39**(suppl 2), W541–W545 (2011)
15. Hartung, M., Groß, A., Rahm, E.: COnto-Diff: generation of complex evolution mappings for life science ontologies. *J. Biomed. Inf.* **46**(1), 15–32 (2013)
16. Doğan, R.I., Leaman, R., Lu, Z.: NCBI disease corpus: a resource for disease name recognition and concept normalization. *J. Biomed. Inf.* **47**, 1–10 (2014)