Journal of Biomedical Informatics 84 (2018) 59–74

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

From lexical regularities to axiomatic patterns for the quality assurance of biomedical terminologies and ontologies

Philip van Damme[a], Manuel Quesada-Martínez[b,c], Ronald Cornet[a], Jesualdo Tomás Fernández-Breis[b,*]

[a] Department of Medical Informatics, Amsterdam Public Health research institute, Academic Medical Center, University of Amsterdam, The Netherlands
[b] Departamento de Informática y Sistemas, Universidad de Murcia, IMIB-Arrixaca, Murcia, Spain
[c] Center of Operations Research (CIO), University Miguel Hernandez of Elche (UMH), Spain

ARTICLE INFO

Keywords:
Ontology quality assurance
Lexical regularities
Axiomatic patterns
SNOMED CT

ABSTRACT

Ontologies and terminologies have been identified as key resources for the achievement of semantic interoperability in biomedical domains. The development of ontologies is performed as a joint work by domain experts and knowledge engineers. The maintenance and auditing of these resources is also the responsibility of such experts, and this is usually a time-consuming, mostly manual task. Manual auditing is impractical and ineffective for most biomedical ontologies, especially for larger ones. An example is SNOMED CT, a key resource in many countries for codifying medical information. SNOMED CT contains more than 300 000 concepts. Consequently its auditing requires the support of automatic methods. Many biomedical ontologies contain natural language content for humans and logical axioms for machines. The 'lexically suggest, logically define' principle means that there should be a relation between what is expressed in natural language and as logical axioms, and that such a relation should be useful for auditing and quality assurance. Besides, the meaning of this principle is that the natural language content for humans could be used to generate the logical axioms for the machines. In this work, we propose a method that combines lexical analysis and clustering techniques to (1) identify regularities in the natural language content of ontologies; (2) cluster, by similarity, labels exhibiting a regularity; (3) extract relevant information from those clusters; and (4) propose logical axioms for each cluster with the support of axiom templates. These logical axioms can then be evaluated with the existing axioms in the ontology to check their correctness and completeness, which are two fundamental objectives in auditing and quality assurance. In this paper, we describe the application of the method to two SNOMED CT modules, a 'congenital' module, obtained using concepts exhibiting the attribute `Occurrence - Congenital`, and a 'chronic' module, using concepts exhibiting the attribute `Clinical course - Chronic`. We obtained a precision and a recall of respectively 75% and 28% for the 'congenital' module, and 64% and 40% for the 'chronic' one. We consider these results to be promising, so our method can contribute to the support of content editors by using automatic methods for assuring the quality of biomedical ontologies and terminologies.

1. Introduction

In recent years, biomedical ontologies and terminologies have been recognised as playing an important role in the achievement of semantic interoperability of clinical information, as reflected in the recommendations of international initiatives such as the FP7 Network of Excellence SemanticHealthNet [1]. The increasing importance of such semantic resources has also stimulated their development and organisation in publicly available repositories. BioPortal [2], which is likely to be the most popular repository of biomedical semantic resources, contains about 700 biomedical ontologies, terminologies and controlled vocabularies.

Ontologies are defined as formal, explicit specifications of shared conceptualisations [3]. The development of semantic resources is usually the result of cooperation between two types of users: domain experts, who provide the domain knowledge, and knowledge engineers, who provide the expertise for the use of semantic formalisms. Ontologies are meant to be useful and processable by both humans and machines. This objective has the implication that the ontology has to include content for both types of intended users. On the one hand,

* Corresponding author.
E-mail addresses: philip.vandamme@student.uva.nl (P. van Damme), mquesada@umh.es (M. Quesada-Martínez), r.cornet@amc.uva.nl (R. Cornet), jfernand@um.es (J.T. Fernández-Breis).

https://doi.org/10.1016/j.jbi.2018.06.008
Received 15 October 2017; Received in revised form 10 June 2018; Accepted 12 June 2018
Available online 14 June 2018

ontologies contain natural language descriptions of their concepts and properties for human consumption. On the other hand, ontologies contain logical axioms, which provide a precise meaning to their concepts and properties when they are expressed in a formal language, for machine consumption.

Generally speaking, the quality of a given product is measured by the degree of fulfilment of the design requirements for such product. The objective of Quality Assurance (QA) processes is to ensure that those requirements are met. This not only includes the identification of errors and making corrections, but also preventing them. The increasing popularity of semantic resources means that more applications are using them, so QA becomes a critical task.

There has actually been an increasing interest in QA and auditing initiatives in recent years [4]. The methodological review presented in [5] proposes a classification based on four criteria: the type of knowledge utilised in the auditing process, the type of techniques used (manual, automated systematic or automated heuristic), the terminology on which the method is focused, the attributes being audited and five quality factors: Concept-orientation; Consistency; Non-redundancy; Soundness; and Comprehensive coverage.

In our current research, we focus on automated systematic methodologies to audit the completeness of concept definitions, which contributes to comprehensive coverage. We propose auditing ontologies by utilising the natural language descriptions associated with concepts, in line with previous studies [6]. Those studies have found that ontologies are richer in natural language content than in logical axioms. The domain knowledge expressed only in natural language is called *hidden semantics* [7]. Concepts in resources such as Gene Ontology (GO) or SNOMED CT have expressive natural language labels because developers tend to use a systematic *naming convention* for the labels of taxonomically related concepts. The use of naming conventions is a principle recommended by the Open Biological and Biomedical Ontology (OBO) Foundry for the construction of ontologies and terminologies. The lexical component of ontologies has already been used for ontology QA in [8], which exploits the semantics associated with the lexical component in ontologies to homogenise the structure of the labels in ontologies. This is done by identifying and transforming labels semantically related but expressed using a different linguistic structure. Hence, the actions taken involve the labels, not the formal concept definitions.

The comparison of what is expressed both logically and in natural language could serve the purpose of QA of biomedical ontologies and terminologies. There should be a correspondence (ideally, an equivalence) between the content expressed in natural language for humans and the content expressed in the form of logical axioms for machines. The lexical content of ontologies such as the GO has been the source of knowledge for natural language processing [9] and has driven the analysis of the compositional structure of GO concepts [10]. In terms of tooling, OBOL [11] facilitates the integration of language and meaning in bio-ontologies, by providing a grammar which permits associating axiomatic patterns with linguistic structures. It was developed for the OBO community and was used for the creation of the GO cross-products [12], and can also be applied for ontology maintenance. In [13], six main types of quality issues in SNOMED CT (see Table 1) were identified. Such issues should be targeted by QA methods. In relation to the incomplete modelling issue, previous works on SNOMED CT [14,15] have identified and illustrated situations where the formal relations are not representing the meaning associated with the natural language content.

Our work is inspired by the 'lexically suggest, logically define' (LSLD) principle [14], which states that the knowledge reflected as natural language in labels should also be represented as logical axioms. Our aim is to design an effective QA method for biomedical semantic resources, which uses resources of natural language content to propose logical axioms. This means that we will mainly address the quality issue of *incomplete modelling* described in Table 1.

In this paper, modules extracted from SNOMED CT, which is the second most audited terminology [5], are used as resources for evaluating the results of the method. Our proposal applies lexical regularities (LRs) (further defined in Section 2.1), which are groups of one or more (consecutive) tokens that appear in several concept labels in an ontology [15,16]. The assumption is that those regularities embed domain knowledge, which should be available as logical axioms. LRs function as seeds for capturing different kinds of issues, which are often concentrated on a group of concepts shared by their textual description. This can be assimilated to the idea of exploiting a 'focus concept' and its neighbourhoods presented in [17]. For example, the SNOMED CT concepts *Pseudocoarctation of aorta* and *Parallel course of aorta and pulmonary artery*, among others, exhibit the LR 'of aorta'. This LR can be used as seed for defining the axiomatic template like `X findingSite some aorta`, which could be applied for all those concepts exhibiting it. The axioms resulting from this process can then be compared with existing axioms to identify missing or incomplete axioms in the ontology. This work contributes to the QA of biomedical ontologies and terminologies by (1) proposing a pattern-based approach, which automatically analyses its lexical content and (2) proposes lexical patterns convertible into axiomatic patterns which can potentially enrich the ontology.

2. Methods

Our QA framework for the extraction of axiomatic patterns from the lexical content in ontologies is graphically described in Fig. 1. The ontology to be analysed is provided as input for the method. The output of the method is a set of axioms extracted from this ontology. The method consists of four main parts:

1. Extraction of LRs from the ontology (Section 2.1).
2. Clustering similar labels from concepts associated with each LR (Section 2.2).
3. Calculation of relevant metrics of the clusters (Section 2.3).
4. Obtaining general axiomatic patterns for each cluster (Section 2.4).

Besides, we also describe the use case (Section 2.5) and propose how to evaluate the effectiveness of the method (Section 2.6).

2.1. Extraction of LRs

The objective of this step is to find and extract the LRs existing in an ontology θ. An ontology θ contains a set of ontology concepts $OC = \{OC_1, ..., OC_n\}$, where n is the number of concepts. For each OC_i, we tokenise and lemmatise [18] its labels obtaining an ordered list of tokens $[T_1, ..., T_m]$, where m is the number of tokens obtained. Conceptually, a label refers to a natural language description associated with a concept in the ontology, which can be represented in the Web Ontology Language (OWL) using the `rdfs:label` annotation property (see the example in Fig. 2). In the case of SNOMED CT, concepts are described in natural language by means of a number of synonyms and one fully specified name, which provides an unambiguous description for a concept by concatenating a description with the name of the semantic tag in brackets, e.g., *Burn scar (morphologic abnormality)* or *Burn scar (disorder)*. In the OWL representation of SNOMED CT, this fully specified name is used for `rdfs:label` annotations. In this paper we use the term 'labels' to refer to the fully specified name of SNOMED CT concepts, without the bracketed name of the semantic tag. In the previous example, both concepts will have the 'label' *Burn scar*.

Conceptually, an LR is a single token (individual word) or a consecutive group of them (multiple words), which appear in several labels of an ontology. The formal definition of an LR is described as:

Definition 1 (*Lexical regularity (LR)*). An ordered sublist of tokens $LRT = [T_1, ..., T_i]$, where $i \in [1, max(m)]$, which is repeated in a subset

Table 1
Quality issues in SNOMED CT found in [13].

Issue	Description	Example
Incorrect schemas	Concepts that are incorrectly classified, regarding the domain knowledge	A problem in the leg is not only classified as a disorder of the lower extremity, but also as a disorder of the abdomen
Misunderstanding of semantics of attributes	Incorrect direct subclass axioms between named concepts	`Hypertension` is classified as both a `Finding` and a `Disorder of Soft Tissue`
Incomplete modelling	Incomplete logical model, no complete definition or missing conditions	`Heart disease` is partially defined as `Disease with some site heart` instead of `Disease with only site heart`
Over-literal definitions	Terms being interpreted too literal, while they have a more specialised meaning	The literal origin of "Neuropathy" can be `Disorder of Nerve`, but the meaning is closer to "dysfunction" of nerves
Not tracing errors to their roots	Finding an error and not tracing the root upwards in the hierarchy	Finding that `Hypertension` is a `Disorder of Soft Tissue`: tracing upwards in the hierarchy shows that the site of `Hypertensive disorder` is some `Artery` and that arteries are `Soft Tissues`, which explains the error
Lack of normalisation	Separating and recombining concepts with appropriate definitions	Using many different aspects to describe related concepts in a branch, for example: '"site", '"stage", "severity", and "symptom" of a disease process

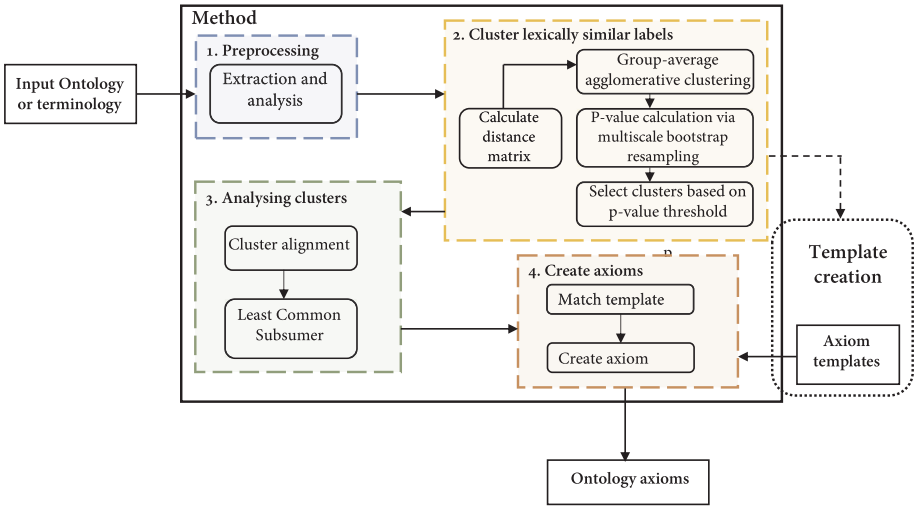

Fig. 1. *Visual model of the method.* The box on the left shows the input. The method is divided into four sequential steps: extraction of LRs (Section 2.1), clustering (Section 2.2), cluster analysis (Section 2.3), and creation of axioms using templates (Section 2.4).

of concepts $OC_{LR} \subset OC$. *Example:* 'of aorta', 'of', and 'aorta' are LRs found in the labels *Rupture of aorta* and *Finding of aorta*.

The lexical analysis has been performed using the OntoEnrich framework [19]. OntoEnrich is implemented in Java and uses the OWL API [20] for processing the ontology and manipulating labels. The Stanford Java NLP API [18] is used for tokenisation and lemmatisation purposes. OntoEnrich has an input parameter, the *coverage threshold*, which enables discarding LRs with the size of OC_{LR} below a certain threshold. Given an ontology θ and a *coverage threshold*, OntoEnrich returns a set of $LRs = LR_1, ..., LR_j$. Each LR_k is defined by: its *LRT* (the tokens in the regularity), its subset of concepts *LROC* (the labels and the concepts which exhibit the regularity).

It should be pointed out that a concept could exhibit more than one LR. Following our running example, 'of aorta' is an LR being exhibited in 41 concepts from a module of SNOMED CT, which contains a total of 18 440 concepts (further details about the module are given in Section 2.5). Fig. 3 (left) shows 9 out of 19 774 LRs found in such a module. Fig. 3 (right) shows 15 concepts which exhibit 'of aorta', including the two concepts mentioned in the introduction. Moreover, LRs of just one token ($m = 1$) are relevant because such repeated fragments in labels could also represent shared domain meaning.

2.2. Clustering LRs

The second step involves processing the labels associated with LRs to obtain clusters of lexically similar labels. For example, we could cluster the 41 concepts which exhibit 'of aorta'.

Definition 2 (*Cluster*). A cluster is a group of lexically similar labels, based on an LR. In this work, the lexical similarity of two labels depends on the number of LRs that are exhibited by both labels. *Example: Rupture of aorta* and *Finding of aorta* belong to the same cluster based on the LR 'of aorta'.

This analysis is performed at the level of LR, so one or more clusters can be proposed for one LR. For this purpose we perform agglomerative hierarchical clustering on the list of labels of one LR, using the group-average linkage algorithm. We apply the Jaccard distance [21] for calculating the distance between two labels (see Formula (1)).

$$d(X, Y) = 1 - \frac{|LR(X) \cap LR(Y)|}{|LR(X) \cup LR(Y)|} \quad (1)$$

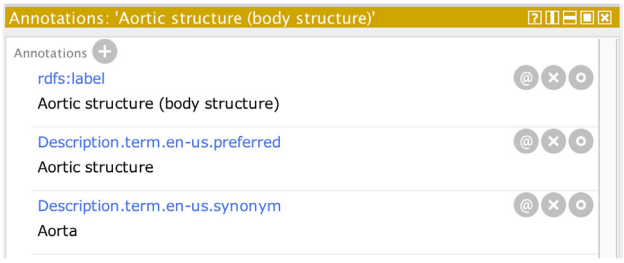

Fig. 2. Example of the three annotations associated with the SNOMED CT concept *Aortic structure (body structure)*.

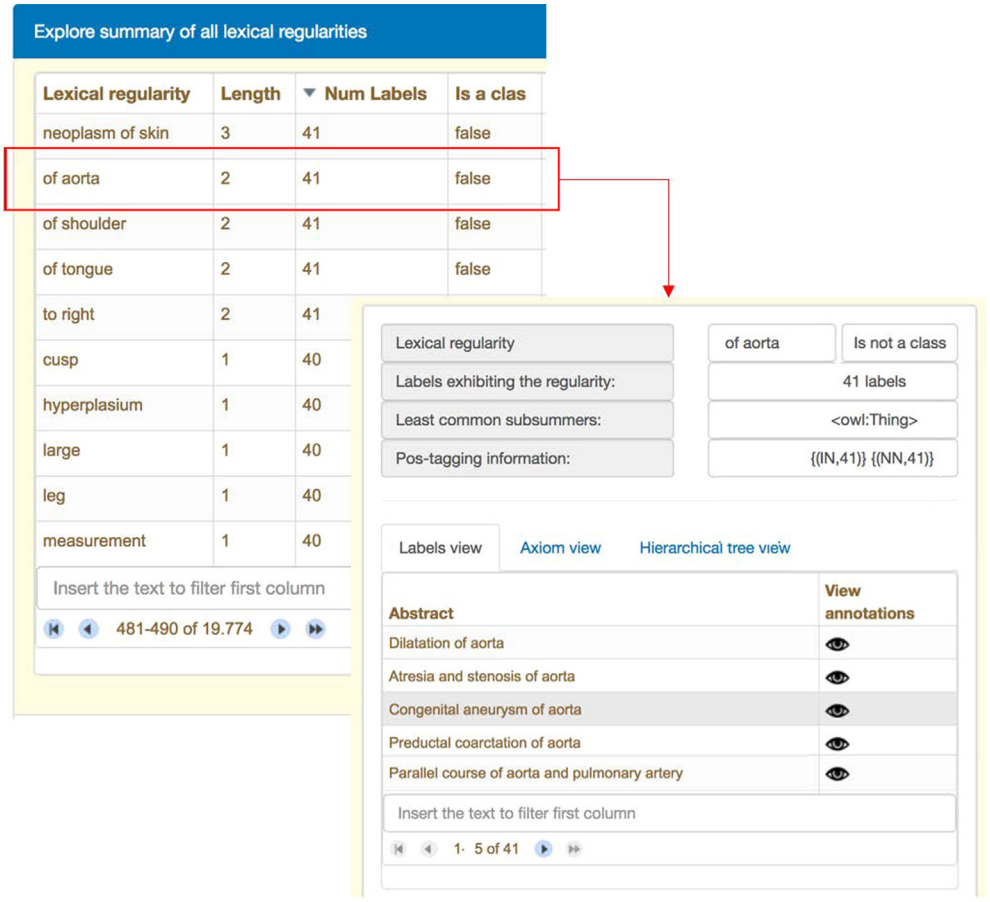

Fig. 3. Screenshots of the OntoEnrich LRs browser.

where $LR(X)$ is the set of LRs that exhibits the label X, and $LR(Y)$ is the set of LRs that exhibits the label Y. Then, $d(X, Y)$ is the distance between the labels X and Y, and it is calculated by using the LRs exhibited by X and Y. The distance $d(X, Y)$ is a number between 0 and 1, where 0 means that the labels X and Y exhibit the same LRs and 1 that the labels do not have any LR in common. The distance matrix contains the distances between all label-pairs. For example, the distance between *Injury of aorta* and *Finding of aorta* is zero if all tokens are found as LRs except 'injury' and 'finding'. In such case both labels exhibit three LRs: 'of aorta', 'of' and 'aorta'.

The result of the clustering is a dendrogram, in which similar labels are located close to each other. Its tree structure enables cutting it at different points, so generating multiple possibilities for the selection of clusters. The R-package `pvclust` [22,23] has been used for the hierarchical clustering. This method calculates probability values (p-values) for each cluster applying bootstrap resampling techniques. Multiscale bootstrap resampling is used for the calculation of the approximately unbiased (AU) p-value and the bootstrap probability (BP) value. The AU p-values and the BP values are shown in Fig. 4 in red (left) and green (right) respectively. We have used 1000 bootstrap replicates, obtaining the corresponding AU p-values and BP values for each cluster. The AU p-value is a better approximation to unbiased p-value than the BP value because it is obtained by multiscale bootstrap resampling and BP is calculated by normal bootstrap resampling.

A cluster AU p-value equal to 0.95 means that this cluster should 'exist' with a significance level of 0.05. Our method selects clusters with an AU p-value equal or greater than 0.95 for further analysis. We actually select the smallest number of clusters with AU p-value $> = 0.95$ that cover the largest number of labels. The labels in clusters with AU p-value lower than the threshold and the non-clustered labels are clustered again once with the objective of obtaining clusters with higher AU p-value by a rearrangement of those labels.

This step can be explained using our running example. Fig. 4 shows the dendrogram associated with clustering sixteen labels exhibiting the LR 'of aorta'. The selected clusters in this example are highlighted by the rectangles ($p > 0.95$), which in this case both share a p-value of 0.98.

2.3. Cluster analysis

A cluster consists of a set of labels that are grouped together according to the number of LRs they share. For example, Fig. 4 shows two clusters consisting of labels exhibiting the LR 'of aorta'. The objective of this step is to provide information about the clusters that could be used for the extraction of axiomatic patterns. For this purpose, two actions are performed on each selected cluster: (1) extraction of a regular expression that represents the labels in the cluster; (2) semantic analysis of the variable part of the regular expression.

2.3.1. Extraction of the cluster alignment

The extraction of the cluster alignment is done by obtaining a lexical alignment of all the labels in a cluster.

Definition 3 (*Cluster alignment*). A cluster alignment is a generalisation that represents the labels in the cluster. This generalisation is based on the lexical similarity between the tokens of the labels. *Example: * of aorta* can be the generalisation of the cluster containing *Rupture of aorta* and *Finding of aorta*.

We have developed an algorithm by adapting the multiple sequence alignment (MSA) of biological sequences [24] to work with our labels. The MSA algorithm is an instance of general edit distance algorithms. The result of the alignment shows how the labels in the cluster can be aligned using their tokens. The result of the alignment is a vector V_i of

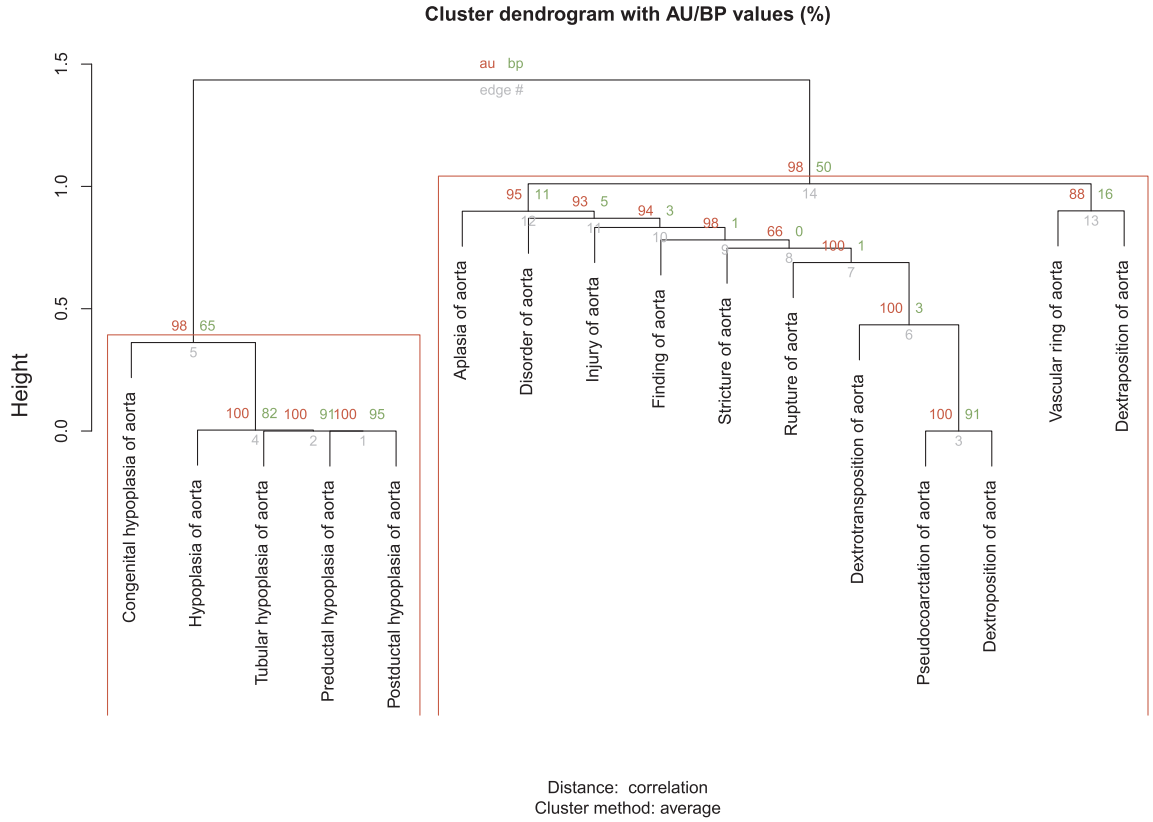

Fig. 4. Dendrogram after clustering part of the labels of the LR 'of aorta'. Two clusters are created shown by the rectangles.

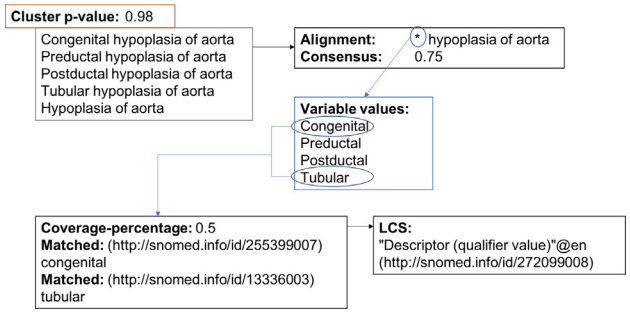

Fig. 5. *Example of the group alignment and least common subsumer.* Results obtained using the LR 'of aorta' for the example shown in Fig. 4.

tokens; each component of the vector contains a set of tokens from different labels which are lexically aligned. For example, the labels *Preductal coarctation of aorta* and *Muscular subvalvar atresia of aorta* contain four and five tokens respectively. Their alignment produces a vector of these five components:

- V_1 = (Preductal, Muscular)
- V_2 = (–, subvalvar)
- V_3 = (coarctation, atresia)
- V_4 = (of, of)
- V_5 = (aorta, aorta).

We say that there is consensus in V_i when all its tokens are equal as in V_4 and V_5. Otherwise, V_i is undetermined (*) or optional (+). V_i is undetermined (*) if it contains at least two different tokens (V_1 and V_3). V_i is optional (+) if it contains at least one gap – (V_2).

The algorithm makes the alignment decision for each position that maximises the global consensus between the labels; the best decision for a given position may be to align two different tokens as happens for V_1 and V_3. This does not mean that tokens in V_1 or V_3 are similar, so this position is considered undetermined. Likewise, gaps are inserted in the position that maximises the alignment score, since they are used for facilitating the consensus between non-consecutive tokens. In this case, the – has been inserted in the first position of V_2, but it would have been inserted in the first position of V_3 if that decision would have produced an alignment with a higher score. In that case, 'coarctation' would have been aligned with 'subvalvar', and 'atresia' with a gap. For more details, we recommend [25]. This algorithm is implemented by adapting MSA algorithms of the BioJava [25] library, and has been included in OntoEnrich.

2.3.2. Semantic analysis of the alignment

A consensus score is calculated for each cluster alignment by obtaining the ratio of tokens with consensus in the alignment. This describes how tokens are aligned between labels, where the identical tokens are associated with the same V_i. Fig. 5 shows the alignment and consensus for one out of six clusters obtained from the LR 'of aorta' shown in Fig. 4.

The alignment has common and variable parts. Following our example, the alignment * *hypoplasia of aorta* has a variable with four different instances: congenital, preductal, postductal and tubular.

The objective of the semantic analysis is to find whether the content of the variable part is semantically related, in this case, whether the four instances are related. The first step is to search for concepts in the ontology being analysed whose label is equal to an instance. In our example, congenital and tubular are found as concepts in SNOMED CT, so having a coverage of 50% (two out of four). There could be alignments with two consecutive variables. In this case the two instances associated with the two variables are joined in a single instance. For example, if 'preductal' and 'coarctation' are instances of two consecutive variables, 'preductal coarctation' would be the instance searched in the ontology.

Fig. 6. SNOMED-CT based figure showing the LCS found for the cluster * *hypoplasia of aorta* from the LR 'of aorta'.

The second step is to find the Least Common Subsumer (LCS) of the concepts associated with the variable instances. The LCS is calculated using the taxonomic relationships defined in the ontology. Fig. 6 shows the LCS and other common ancestors of the instantiations found for the * *hypoplasia of aorta* cluster. In this figure, the LCS `Descriptor (qualifier value)` shows the relationship between the concepts 'tubular' and 'congenital'. It should be noted that one or more LCSs can be found. This method has been implemented in Java and included in OntoEnrich.

2.4. Creation of axioms using templates

The last step of the method is the extraction of the axiomatic templates and the creation of axioms for the concepts. This step uses the cluster alignments to create axiomatic patterns for each cluster, and requires two actions: (1) creation of templates; and (2) creation of the axiomatic patterns. The implementation of the methods described in this section has been done in Java and R.

OWL ontologies can be seen as consisting of a set of logical axioms that provide the description of concepts, properties and constraints. There are two basic types of axioms in OWL: (1) *subClassOf*, which permits definition of hierarchical relations between concepts, establishing necessary conditions for class membership; and (2) *equivalentClass*, which allows definition of equivalence relations between concepts, establishing sufficient conditions for class membership. Let us illustrate OWL modelling with SNOMED CT concepts and attributes. Fig. 7 shows the definition of the concept *Congenital hypoplasia of aorta* both in the SNOMED CT browser [26] and in the Protégé ontology tool.[1] The relationship between the concepts *Congenital hypoplasia of aorta* and *Aortic structure* is represented by the attribute `Finding site` (see the left square of the upper part of Fig. 7, number 1), which is represented in OWL as part of an *equivalentClass* axiom.

OWL ontologies have two different models: asserted and inferred. The asserted model includes all the axioms that have been explicitly defined in the ontology. The asserted axioms (see content inside the dotted squares in Fig. 7) are used by a reasoner to deduce new axioms, so obtaining the inferred model. The role of these models is similar to the stated and inferred ones in SNOMED CT. Examples of inferred axioms are shown in Fig. 7: *Congenital hypoplasia of aorta* is a subclass of *Hypoplasia of aorta* and a subclass of *Congenital anomaly of aorta*. The inferred model of the ontology contains all the knowledge, so this is the model used by our method to obtain the LCS.

2.4.1. Creation of the templates

A training set of cluster alignments is used to define the templates. The templates are defined by examining the alignments manually. It should be noted that a template can be matched with the alignment of many clusters and that we aim at designing general, reusable templates.

For example, let us consider the alignment * *of aorta*. By representing '*' as `[variable(s)]`, 'of' as `[preposition]` and 'aorta' as `[something]`, the resulting template would be `[variable(s)] [preposition] [something]`. Moreover, during the definition of the templates, they are associated with an OWL axiomatic pattern. For

[1] http://protege.stanford.edu/.

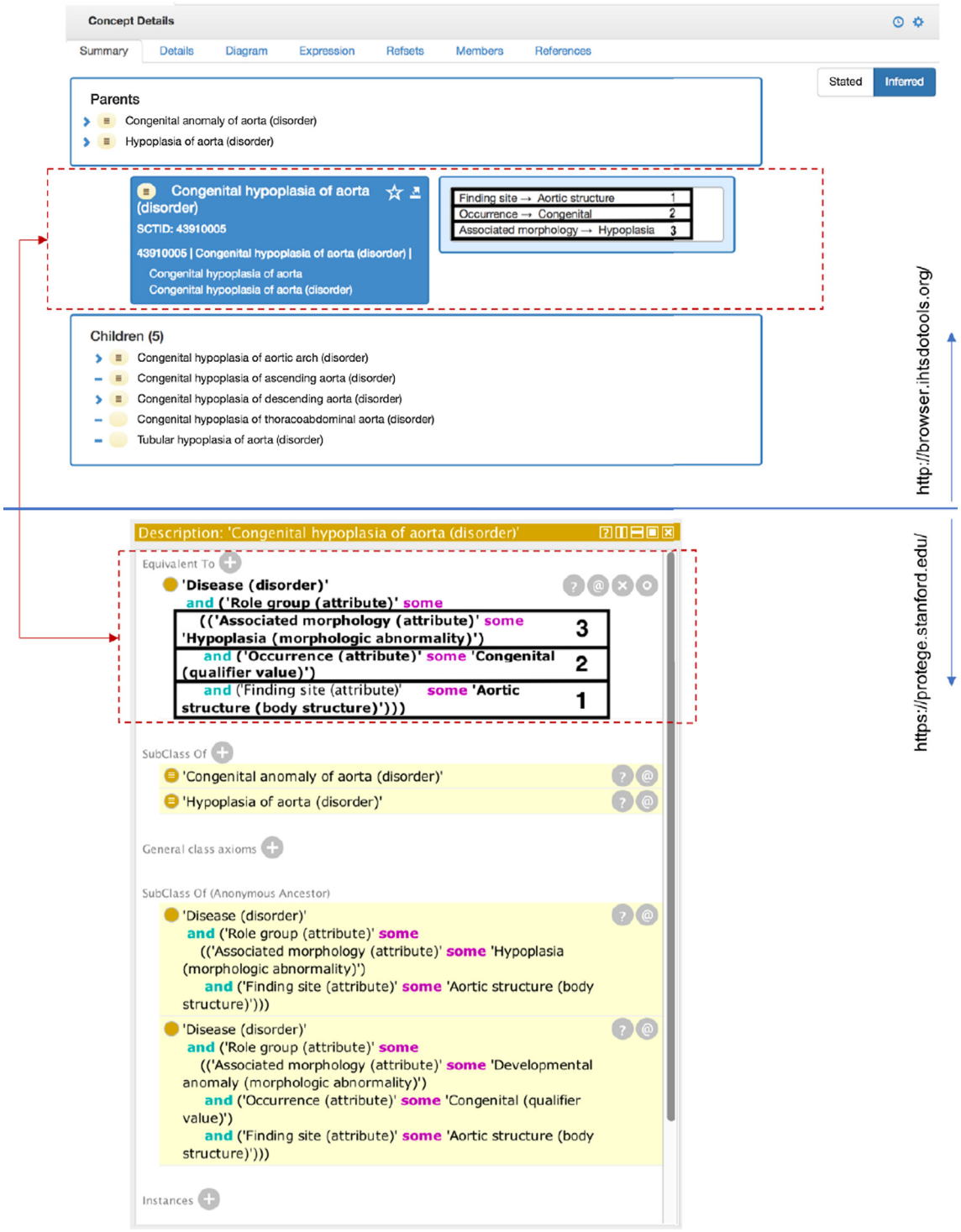

Fig. 7. Representation of the concept `Congenital hypoplasia of aorta` in the official SNOMED CT browser, and using OWL axioms in Protégé.

instance, the template `[variable(s)] [preposition] [something]` may be associated with the OWL pattern `[something] AND hasProperty(preposition) SOME [variable(s)]`. This represents an axiom stating that `[something]` is associated with the `[variables(s)]` through a property which depends on the preposition. For example, considering the axioms already defined in SNOMED CT (see Fig. 7), the preposition 'of' may be linked with the property `Finding site (attribute)` so this axiomatic pattern would create the axiom `Finding site (attribute) some Aortic structure (body structure)` for all concepts included in a cluster matching this template.

2.4.2. Creation of axioms

The templates are used for creating general axioms, which can be applied to a group of lexically similar concepts. The method needs to use a set of preposition-property associations to complete the axiomatic patterns.

Let us consider the LR 'of aorta'. Labels exhibiting this regularity are classified in clusters like * * *of aorta*, * *hypoplasia of aorta*, * + *atresia of aorta* and so on. The cluster * * *of aorta* matches our running template

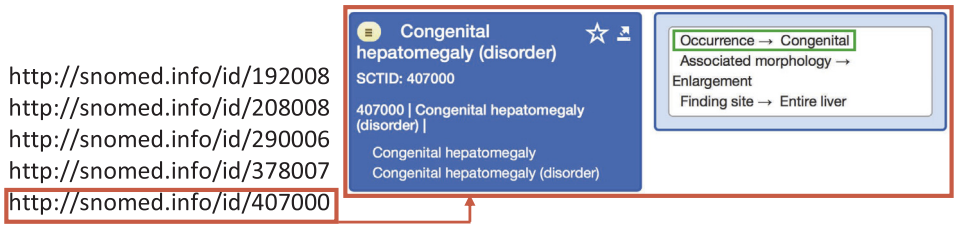

Fig. 8. *Fragment of the first five concepts of the seed signature used for obtaining the congenital module.* The seed signature consists of a list of SNOMED CT concepts, which in this case are descendants of `Disease (disorder)` and exhibit the attribute `Occurrence – Congenital`.

`[variable(s)] [preposition] [something]`. If our template links the preposition 'of' with the property `Finding site (attribute)`, the axiomatic pattern created for this cluster would be `* * AND 'Finding site (attribute)' SOME aorta`.

It should be noted that the 'property value' (e.g. 'Finding site') included in the pattern is associated with the matched string using our lexical alignment algorithms. For example, when the cluster * * *of aorta* is found, the template matches the preposition of the pattern, so the algorithm can deal with the variables '* *' stated before 'of' and the consensus 'aorta' after it. According to Fig. 7, there is no direct relation between 'aorta' and the concept *Aortic structure (body structure)* through `rdfs:label`. As we mentioned in Section 2.1, the `rdfs:label` property of SNOMED CT concepts is the fully specified name, which includes the bracketed name of the semantic tag. However, our lexical alignment algorithm does not use the name of the semantic tag. Moreover, SNOMED CT concepts are also associated with synonyms, which are also taken into account by our lexical alignment method. This algorithm is based on lexical techniques such as tokenisation and lemmatisation, which allow us to align the string 'duct' and 'ducts' (singular-plural), or to detect that the string 'aorta' is semantically represented by the concept *Aortic structure (body structure)*.

Finally, if 100% of the variable instances are concepts in the ontology, the LCS would replace the variables in the general axiomatic pattern (for one group of clustered labels). Otherwise, the LCS information of the instances found can be shown to domain experts for their manual inspection to help them to define more and better templates. In our running example, the cluster * *hypoplasia of aorta* (see Figs. 5 and 6) provides the following insight to domain experts. On the one hand, the LCS of 50% of the concepts are classified in this cluster as 'Descriptor (qualifier value)', concretely 'congenital' and 'tubular'. On the other hand, 'preductal' (related to the part of the aorta proximal to the aortic opening of the arterial canal) and 'postductal' (related to that part of the aorta distal to the aortic opening of the arterial canal) were not found as concepts in SNOMED CT. In case they were included as qualifier values, the LCS could be automatically applied in the template.

2.5. The case study

We have applied the method to two modules in the SNOMED CT International Release of January 2015. As this release of SNOMED CT contains 311 532 concepts and the validation will require manual effort, we have applied our method to smaller subsets of SNOMED CT.

In order to obtain such subsets, we used mechanisms for the automatic extraction of ontology modules [27,28]. In particular, we used locality-based modules [29]. A locality-based module M is a subset of the axioms in an ontology θ, which is extracted from θ for a set S of concepts (concept or property names). The set S is called a *seed signature* of M. Informally, everything the ontology θ can infer about the topic consisting of the concepts in S and M, is already known by its module M. Further information about how the extraction of the module is operationalised can be found in [29] (section 'How do locality-based modules work?'). Here, we used the following modules:

- Module *congenital*: This module is based on a potential modelling issue identified by SNOMED International in the SNOMED CT January 2015 release and linked to the axiomatic representation of concepts: 'artf229197: Congenital Occurrence vs. congenital Morphology'.[2] This was the main reason for working with the January 2015 release. We generated a seed signature that contained all the concepts which are asserted descendants of `Disease (disorder)` and which also exhibit the attribute `Occurrence – Congenital`. In order to achieve this, we developed an algorithm which was implemented in Java using the OWL API. This algorithm analyses the axioms in the asserted model of the ontology. As a result, our obtained seed signature contains 6600 disorders. Fig. 8 shows a fragment of the first five concepts of this seed signature including an example of one concept in SNOMED CT. Afterwards we used the module extractor [29] using this signature as input, and we obtained a module which contains a total of 18 440 SNOMED CT concepts.
- Module *chronic*: This module is based on previous work that focused on the analysis of syntactic regularities and irregularities in SNOMED CT [30]. This work analysed concepts which have the word 'chronic' expressed in their labels. We used this as a reference for creating a seed signature containing all the asserted descendant concepts of `Disease (disorder)` which exhibit the `Clinical course – Chronic` attribute, using the algorithm mentioned above. As a result, the *seed signature* contained 165 diseases. The extracted module contains a total of 3262 SNOMED CT concepts.

We consider these two modules of interest for several reasons. Firstly, they allow us to apply and validate our method with two simplified versions of SNOMED CT, and in particular contexts already mentioned in earlier works regarding the QA of SNOMED CT. Secondly, we combine ideas used in [27,28,31] for keeping all the logical axioms in the module related to the seed signature. Thus, the modules represent the whole version from a logical point of view. Moreover, use of smaller modules is especially relevant for our method because it involves manual evaluation of the created axioms.

In this experiment, the LRs have been extracted by processing the fully specified names (without the bracketed name of the semantic tag) of SNOMED CT concepts. Fig. 3 shows part of the information offered by OntoEnrich for the LRs found in the SNOMED CT congenital module. The length and number of labels that exhibit the LR and whether the regularity is a concept in SNOMED CT is shown for each regularity. Some labels exhibiting the regularity 'structure of left' are shown on the right side. Only LRs with frequency (or coverage) equal to or greater than 0.01% of the number of concepts in the modules and exhibited by at least six labels have been further processed. Both parameters are thresholds. The first one is set based on an experimental decision. The higher the frequency value, the more specific the regularities we obtain. In this experiment, we are interested in obtaining as many LRs as possible, so we use a low coverage value. The threshold of six labels is a limitation of the clustering library used. No additional filters have been

[2] http://sele.inf.um.es/ontoenrich/projects/axiomatic_patterns/cluster_lrs_2017/files/artfl229197.png.

applied. Twenty LRs have been randomly selected in each module for performing the evaluation.

2.6. Evaluation

The main goal of the evaluation is to analyse the relation between the axioms proposed by our method and the axioms included in the OWL version of SNOMED CT. This will permit to study to what extent the method is able to reconstruct existing axioms and to propose potentially missing ones. The evaluation results will be analysed using four categories:

- True positive (TP): The axiomatic pattern created by the method is also codified in the ontology by axioms.
- False positive (FP): The axiomatic pattern created by the method is not codified in the ontology by axioms. From a QA perspective, FPs represent potentially missing axioms.
- True negative (TN): The method did not create an axiomatic pattern, and no axioms representing the pattern were found in the ontology either.
- False negative (FN): A set of axioms representing the axiomatic pattern was found in the ontology, but the method did not create them.

The axioms of the inferred model of SNOMED CT, which are obtained using the Snorocket reasoner [32], are the gold standard for the evaluation. The classification of the axioms suggested by our method in TP or FP is the result of the manual comparison of these axioms with the axioms associated with the SNOMED CT concepts in the gold standard. To facilitate this evaluation, the properties used for creating our axioms were selected from the ones used in SNOMED CT. Thus, we imposed the condition of equality to consider our axioms as TP. Using the asserted model instead of the inferred one could lead to suboptimal evaluation results. The asserted model has fewer axioms than the inferred one, which means that some TPs in the inferred model could be FP with the asserted model. Finally, precision (Formula (2)), recall (Formula (3)) and F-measure (Formula (4)) have been calculated.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F\text{-}measure = \frac{2}{\frac{1}{precision} + \frac{1}{recall}} \tag{4}$$

3. Results

In this section we present a summary of the results obtained in our case study. The full seed signatures, lexical analysis and method output files of both modules can be found at our website.[3]

3.1. Axiom templates

A training set of cluster alignments has been used to define the axiom templates for this validation. The defined templates are shown in Table 2, including the matching axiomatic patterns and examples. Templates 1 and 2 include prepositions, which are associated with specific properties as follows:

- Preposition 'of' - Property `Finding site (attribute)`
- Preposition 'due to' - Property `Due to (attribute)`

[3] http://sele.inf.um.es/ontoenrich/projects/axiomatic_patterns/cluster_lrs_2017.

Table 2
The axiom templates defined in this experiment.

Template	Axiomatic pattern	Example
1. [VARIABLE(S)] [PREPOSITION] [something]	[VARIABLE(S)] [PROPERTY VALUE] some [something]	Alignment: * *of aorta* Axiom: * `AND Finding site SOME aorta`
2. [something] [PREPOSITION] [VARIABLE(S)]	[something] [PROPERTY VALUE] some [VARIABLE (S)]	Alignment: *rupture of aorta due to* * Axiom: `rupture of aorta Due to SOME *`
3. [VARIABLE(S)] [something]	[VARIABLE(S)] subClassOf [something] (only if 'something' is found as a concept)	Alignment: + *dilatation of aorta* Axiom: `+ subClassOf dilatation of aorta`

3.2. Results on the SNOMED CT congenital module

3.2.1. General description

We determined 19774 LRs for the congenital module. Table 3 shows the results of processing the 20 randomly selected LRs and their properties. For each regularity the table shows whether it is a concept in the module (two are concepts), the number of labels in which the regularity is exhibited (the mean is 26), the number of clusters created by the method, 71 in total, and the percentage of labels not included in the selected clusters based on the p-value threshold (0.95). All the labels of 11 LRs are included in a selected cluster. For nine LRs a percentage of labels between 21% and 100% are not included in a selected cluster. No cluster has been selected for the LR 'recessive muscular'. The average cluster p-value, the median alignment consensus and the median percentage of variable instances found as a SNOMED CT concept are provided for the clusters of a regularity. The last column shows the number of axiomatic patterns created for each regularity (one pattern is created for a cluster), 20 in total (out of 71 clusters).

We illustrate part of the output generated by the method for the LR 'of aorta' (see Fig. 9). The left and right parts of the figure show the two clusters obtained for this LR, whose p-values exceed the threshold 0.95. Both clusters correspond to the clusters shown earlier in the dendrogram of Fig. 4. The cluster on the left includes five labels, whose alignment is * *hypoplasia of aorta*. The axiomatic pattern * *'subClassOf' hypoplasia of aorta* was created. The cluster on the right includes 11 labels whose alignment is * * *of aorta*. The axiomatic pattern created is: * * *'Finding site (attribute)' some aorta*.

3.2.2. Results by axiomatic patterns

Table 4 shows the results of the evaluation of the axiomatic patterns extracted for the clusters. The table contains the TP, TN, FP and FN for each LR at both the level of clusters and the level of labels. The results at the level of cluster reveal a precision of 75% and a recall of 28%. Similar results are obtained at the level of labels: precision of 72% and a recall of 29%. The F-measure for both levels is, respectively, 0.41 and 0.41. Out of the 20 axiomatic patterns created 15 are also present in SNOMED CT (TP) and five are not (FP). There is also no axiomatic pattern present in SNOMED CT for 13 clusters for which the method has not extracted any (TN). Thirty-eight clusters for which the method has not extracted any axiomatic pattern have axiomatic patterns in SNOMED CT (FN).

3.2.3. Results by templates

Table 5 shows the results per template. Template 1 is matched by nine axiomatic patterns: seven TP and two FP; template 2 is matched by

Table 3
Results of the application of the method to the 20 LRs of the congenital module.

LR	Is a concept?	# Labels	# Clusters	Not-clustered labels (%)	Cluster p-value average	Alignment consensus (clusters median %)	Variable instances found in SNOMED CT (clusters median %)	# Axiomatic patterns created
– baby	No	10	2	40%	0.9819	54%	0%	0
bronchopulmonary	No	6	1	67%	0.9999	67%	0%	0
chromosomal	No	15	2	0%	0.9671	29%	86%	0
disease due	No	23	2	0%	0.9724	51%	62%	1
divide left atrium with all pulmonary vein	No	7	2	0%	0.9785	61%	0%	0
duct	No	233	26	9%	0.9779	31%	25%	7
ectropion	Yes	7	2	0%	0.9793	33%	50%	0
Epidermolysis bullosa	Yes	8	3	0%	0.9926	67%	33%	0
mixed	No	12	1	33%	0.9622	8%	50%	0
of aorta	No	41	8	29%	0.9886	68%	50%	5
of cervix	No	34	2	0%	0.9639	20%	100%	0
of subclavian	No	7	2	0%	0.9792	78%	67%	1
operative procedure	No	49	3	37%	0.9823	63%	50%	1
posterior segment of eye	No	6	2	0%	0.9776	79%	50%	2
recessive muscular	No	6	0	100%	–	–	–	0
red blood	No	14	4	21%	0.9790	67%	50%	1
segment of eye	No	10	2	0%	0.9836	51%	50%	1
sensorineural	No	6	2	0%	0.9866	41%	83%	1
sensory	No	15	3	60%	0.9789	50%	0%	0
septum with	No	6	2	0%	0.9814	81%	25%	0
							Total:	20

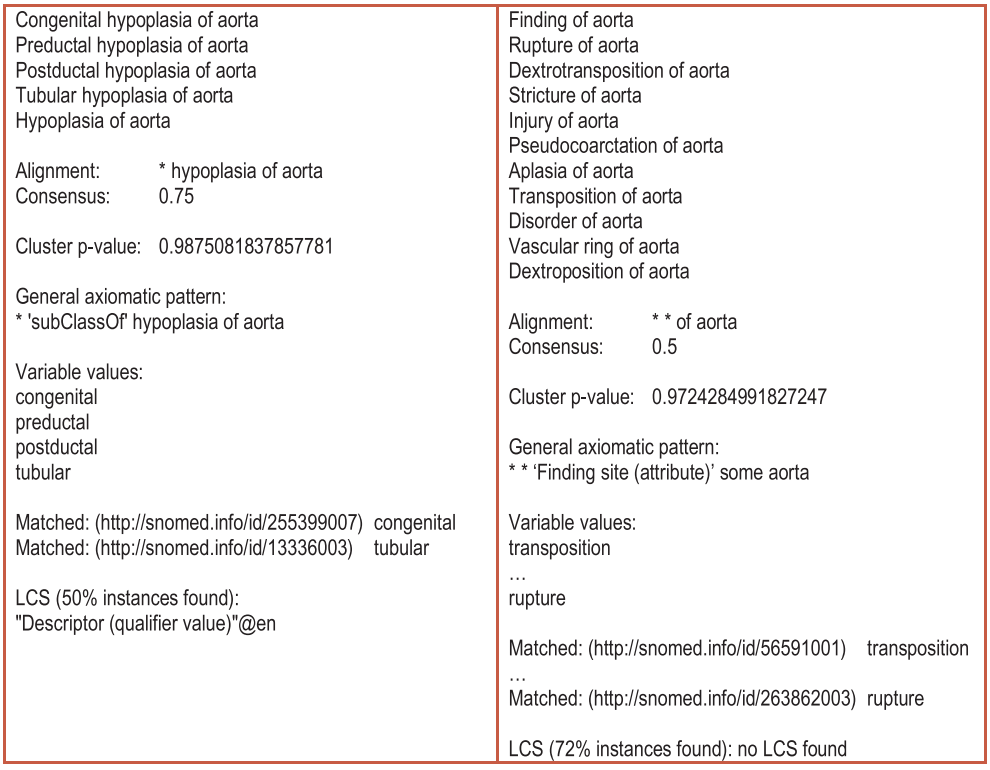

Fig. 9. *Partial output of the LR 'of aorta'.* Two clusters are shown (left and right), corresponding to the cluster dendrogram of Fig. 4.

three axiomatic patterns: one TP and two FP; template 3 is matched eight times: seven TP. In Fig. 9, the left and right hand side show cases of respectively FN and TP clusters of our running example. As an example, Fig. 10 shows the axioms used by the concept *Finding of aorta* and its relationship with the axiomatic pattern proposed by our method; no LCS has been included in the axiom because not all variable instances were found as SNOMED CT concepts. Fig. 11 shows an example of a TP cluster found for the LR 'duct', the alignment and the axiomatic pattern (blue rectangle) generated by the method. The axiom information from Protégé is shown for the red marked label. The axiomatic pattern created matches the actual SNOMED CT axiom for all four labels. The LCS is included in the axiom, and all variable instances are SNOMED CT concepts. Fig. 12 shows an FP cluster found for the LR 'disease due', the alignment of the cluster, the axiomatic pattern and the LCS of 54% of the variable instances. The axiom information from Protégé for the red marked label is shown. It is a false positive because the used property value `Due to (attribute)` is not the correct one. The correct property value is `Causative agent (attribute)`, a sibling of `Due to (attribute)`.

Table 4
Results of the evaluation of the axiomatic patterns for the congenital module.

LR	Per cluster				Per label			
	# TP	# TN	# FP	# FN	# TP	# TN	# FP	# FN
– baby	0	0	0	2	0	0	0	6
bronchopulmonary	0	0	0	1	0	0	0	2
chromosomal	0	1	0	1	0	8	0	7
disease due	0	1	1	0	0	10	13	0
divide left atrium with all pulmonary vein	0	0	0	2	0	0	0	7
duct	6	7	1	12	40	149	3	33
ectropion	0	0	0	2	0	0	0	7
Epidermolysis bullosa	0	0	0	3	0	0	0	8
mixed	0	1	0	0	0	8	0	0
of aorta	4	0	1	3	18	0	5	6
of cervix	0	0	0	2	0	0	0	34
of subclavian	0	0	1	1	0	0	3	4
operative procedure	0	0	1	2	0	0	2	29
posterior segment of eye	2	0	0	0	6	0	0	0
recessive muscular	0	0	0	0	0	0	0	0
red blood	1	2	0	1	2	6	0	3
segment of eye	1	0	0	1	4	0	0	6
sensorineural	1	0	0	1	4	0	0	2
sensory	0	1	0	2	0	2	0	4
septum with	0	0	0	2	0	0	0	6
Total:	15	13	5	38	68	183	26	164
	Precision:	75%	Recall:	28%	Precision:	72%	Recall:	29%

Table 5
Results per template obtained for the congenital module.

Template	# Total matches	# TP	# FP
1. [VARIABLE(S)] [PREPOSITION] [something]	9	7	2
2. [something] [PREPOSITION] [VARIABLE (S)]	3	1	2
3. [VARIABLE(S)] [something]	8	7	1

3.3. Results for the SNOMED CT chronic module

3.3.1. General description

We found 2783 LRs for the chronic module. Table 6 shows the results of 20 randomly selected LRs for the chronic module. Four LRs were found as concepts in this module. The mean number of labels per LR is 19. For 16 LRs a percentage between 9% and 63% of the labels has not been included in any cluster. Eighty-six clusters were generated by our method, and a general axiomatic pattern was created for 36 clusters. At least one cluster has been created for every selected LR.

3.3.2. Results by axiomatic patterns

Table 7 shows the results of the evaluation of the axiomatic patterns created for the chronic module. At the cluster level, a precision of 64% and a recall of 40% have been obtained. At the label level, the precision and recall are respectively 50% and 30%. The method resulted in 36 axiomatic patterns, of which 23 were present in SNOMED CT (TP) and 13 were not (FP). No axiomatic pattern was created for six clusters by the method, and for those clusters there was no axiomatic pattern present in SNOMED CT either (TN). No axiomatic pattern was created for 35 clusters while there was an actual axiomatic pattern present in SNOMED CT (FN). The F-measure for respectively the cluster and label levels is 0.49 and 0.38.

3.3.3. Results by templates

Table 8 shows the results per template. Template 1 is matched by five axiomatic patterns: two TP and three FP; template 2 is matched by eight axiomatic patterns: four TP and four FP; template 3 is matched 24 times: seventeen TP and seven FP.

3.4. Precision and recall per template

Table 9 shows the precision and recall results per module when systematically removing one template, as well as the results with only a single template. The first row contains the overall precision and recall for both modules as shown before.

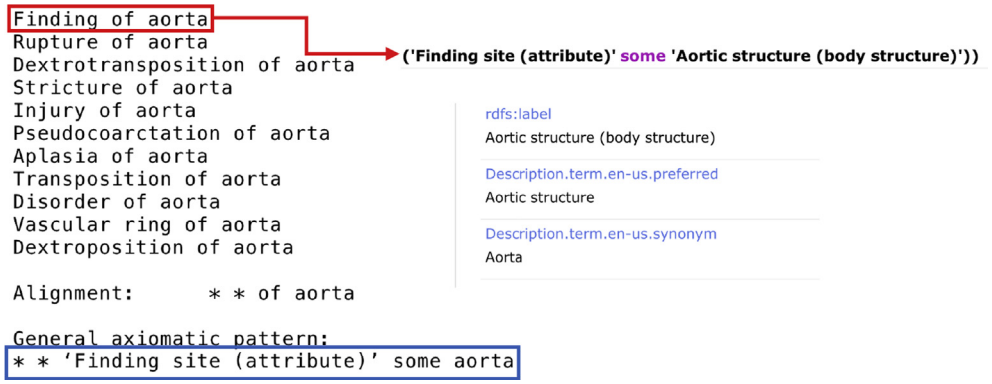

Fig. 10. True positive example for the LR 'of aorta'. The least common subsumer has not been included because not all variable instances has been found as a concept in SNOMED CT. The label `Aortic structure (body structure)` is also defined by the synonym "Aorta", which has been used by our algorithm in the general axiom.

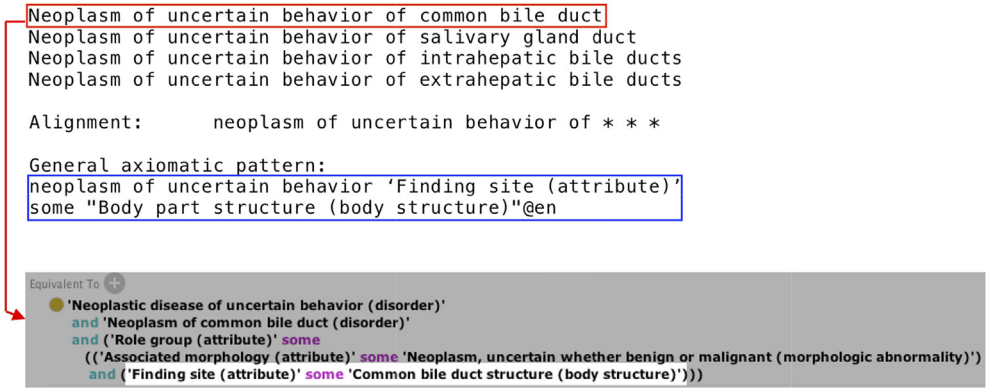

Fig. 11. True positive example for the LR 'duct'. Here the least common subsumer has been automatically included in the axiomatic pattern.

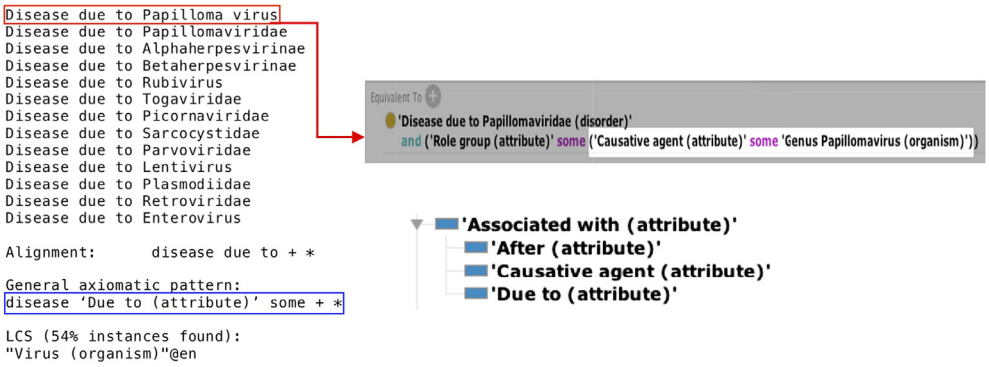

Fig. 12. False positive example for the LR 'disease due'.

Table 6
Results of the application of the method to 20 LRs of the chronic module.

LR	Is a concept?	# Labels	# Clusters	Not-clustered labels (%)	Cluster p-value average	Alignment consensus (clusters median %)	Variable instances found in SNOMED CT (clusters median %)	# Axiomatic patterns created
aneurysm of	No	11	3	45%	0.9674	60%	50%	2
arterial	No	15	5	33%	0.9735	67%	100%	4
chronic kidney disease stage	No	10	3	40%	0.9839	63%	0%	0
colon	No	17	3	29%	0.9906	67%	100%	3
deep venous thrombosis of	No	13	4	0%	0.9732	74%	92%	3
enterovirus	No	6	1	17%	0.9766	17%	67%	0
hereditary	No	21	3	24%	0.9764	33%	0%	1
infection of the central	Yes	8	1	63%	0.9993	75%	67%	0
insufficiency	Yes	8	3	13%	0.9972	80%	50%	2
obstruction	No	33	8	36%	0.9696	58%	50%	4
of substance	No	6	2	0%	0.9899	37%	33%	0
operation on	No	77	14	21%	0.9839	46%	74%	0
operative	No	24	4	50%	0.9864	53%	56%	1
pneumonitis	Yes	8	3	0%	0.9760	43%	50%	2
purpura	Yes	11	2	9%	0.9620	25%	63%	1
streptococcal	No	10	2	40%	0.9986	54%	100%	1
tissue sample	No	10	3	30%	0.9786	67%	67%	0
tree	No	6	2	0%	0.9937	58%	50%	1
vascular structure	No	17	4	35%	0.9802	63%	65%	3
vein	No	72	16	22%	0.9810	71%	25%	8
							Total:	36

4. Discussion

4.1. The method

The QA of biomedical ontologies and terminologies is a fundamental task to ensure that the semantic artefacts used in biomedical applications are not the cause of unexpected, wrong behaviours and results. In this work, QA has been approached from the perspective of analysing the relation between the content expressed in natural language for humans and the content expressed in the form of logical axioms for the machines.

Our method pursues to create axioms from the existing LRs in

Table 7
Results of the evaluation of the axiomatic patterns for the chronic module.

LR	Per cluster				Per label			
	# TP	# TN	# FP	# FN	# TP	# TN	# FP	# FN
aneurysm of	1	0	1	1	2	0	2	2
arterial	3	0	1	1	6	0	2	2
chronic kidney disease stage	0	0	0	3	0	0	0	6
colon	2	0	1	0	4	0	8	0
deep venous thrombosis of	3	0	0	1	10	0	0	3
enterovirus	0	0	0	1	0	0	0	5
hereditary	0	1	1	1	0	3	8	5
infection of the central	0	0	0	1	0	0	0	3
insufficiency	1	0	1	1	2	0	3	2
obstruction	3	0	1	4	7	0	4	10
of substance	0	1	0	1	0	3	0	3
operation on	0	0	0	8	0	0	0	61
operative	0	0	1	2	0	0	2	10
pneumonitis	2	0	0	1	6	0	0	2
purpura	1	1	0	0	6	4	0	0
streptococcal	1	0	0	1	2	0	0	4
tissue sample	0	0	0	3	0	0	0	7
tree	1	1	0	0	3	3	0	0
vascular structure	1	0	2	1	5	0	4	2
vein	4	2	4	4	8	4	29	15
Total:	23	6	13	35	61	17	62	142
	Precision:	64%	Recall:	40%	Precision:	50%	Recall:	30%

Table 8
Results per template obtained for the chronic module.

Template	# Total matches	# TP	# FP
1. [VARIABLE(S)] [PREPOSITION] [something]	5	2	3
2. [something] [PREPOSITION] [VARIABLE (S)]	8	4	4
3. [VARIABLE(S)] [something]	24	17	7

Table 9
Precision and recall when systematically removing one template, including the results with only a single of the three defined templates.

Template number(s)	Congenital		Chronic	
	Precision	Recall	Precision	Recall
1 + 2 + 3	75%	28%	64%	40%
1 + 2	67%	17%	46%	14%
1 + 3	82%	27%	66%	35%
2 + 3	73%	17%	66%	38%
1	78%	15%	40%	5%
2	33%	2%	50%	10%
3	88%	16%	71%	33%

ontologies, and our working hypothesis has been that lexically similar labels should also be axiomatically similar by the application of the LSLD principle. This work demonstrates that the development of semi-automatic methods for supporting ontology and terminology content editors in the auditing of the ontology is possible. For this purpose, we have been able to use existing tools such as OntoEnrich to extract and obtain relevant information about the LRs in the semantic resources and hierarchical clustering techniques have contributed to analyse, group and select which regularities are more suitable for the generation of axiomatic patterns.

The content of LRs could correspond to the full label of concepts or object properties in the ontology. Correspondence of an LR with a concept could make the method assume that all the concepts exhibiting the LR should be linked to that concept. If the LR corresponds to the full label of an object property, then the method could assume that all the concepts exhibiting the LR should use such property.

The method has been applied to two modules extracted from SNOMED CT. Semantic resources usually group related concepts in modules. One way of implementing modularity in ontologies is by using subhierarchies, one per module. For example, SNOMED CT is organised in 19 modules such as body structure, finding, event, observable entity, and qualifier value. Nevertheless, there are other ways for defining modules in ontologies, such as the locality-based approach applied in this work. Our two modules do not correspond to any of the 19 SNOMED CT modules, and they cover around 7% of the concepts included in SNOMED CT. The locality-based approach is a flexible way to define semantically-related modules over a large ontology such as SNOMED CT.

In this work we have used hierarchical clustering to analyse the LRs to identify the ones suitable for the creation of axiomatic patterns. This unsupervised technique has been used to require no prior knowledge about the labels and the number of clusters. The quality and usefulness of the clusters obtained can be evaluated by the classification of the axiomatic patterns associated with them as TP, FN, FP or TN. Clusters with TP and FN are the most useful ones because the concepts associated with the labels in the cluster share axioms. An FP cluster might be useful if further examination reveals that the axiomatic pattern created by the method should be included in the semantic resource. The TN classification of a cluster may be due to the lack of templates, so a TN cluster might be turned into FP if more templates are available. According to our results, the clustering technique has demonstrated its usefulness for supporting the process of obtaining the axiomatic patterns. In the SNOMED CT congenital module, 75% (53 out of 71) of the clusters contain labels whose concepts have axioms in common (15 TP and 38 FN). No axioms in common were found for the remaining 18 clusters (13 TN and 5 FP) in SNOMED CT. In the SNOMED CT chronic module 67% (58 out of 86) of the clusters contain labels whose concepts have axioms in common (23 TP and 35 FN).

FN clusters are of interest for QA because they contain labels which share at least one logical axiom, which was not created by our method. The reason could be: (1) a lexical suggestion for this axiom was not detected due to the lack of templates, or (2) the shared axiom was not lexically suggested by the labels. The first case would be solved by providing an appropriate template. The second one is out of the scope

of this work.

The quality and usefulness of the patterns can then be evaluated in terms of precision and recall for each module. The results for the SNOMED CT congenital module reveal a precision of 75% and a recall of 28% (Table 4, cluster level). There is an axiomatic pattern in SNOMED CT for 38 clusters that was not created by the method (FN, Table 4). The results for the SNOMED CT 'chronic' module reveal a precision of 64% and a recall of 40% (Table 7, cluster level).

FN rates have an impact on recall. A low recall in our method means that it is not able to propose axioms already existing in SNOMED CT. Our method creates the axioms based on the templates used, so the number and quality of the templates play a fundamental role in the recall score. In the validation experiment we have used three different templates (Table 2), including two types of property values from SNOMED CT linked to prepositions. The results of our experiments (see Tables 5 and 8) show that the axiomatic patterns of each template were at least one or more times found as actual axiomatic patterns in SNOMED CT (TP). Table 9 shows that the recall is in general lower when fewer templates are used.

The results are promising, taking into account the number of templates used in the experiments. The quality of the results obtained is influenced by the number and types of templates created. Having more templates available would help to reduce the FN rate and, therefore, to increase the TP rate and recall. However, recall will also be limited by the axioms that might not be learnt by just analysing the content of the labels.

4.2. Practical impact and application to other ontologies

Among the SNOMED CT quality issues shown in Table 1, our method contributes especially to *incomplete modelling* since the natural QA application of our method is to propose a set of axioms that should exist in the ontology. We believe that the more complete and precise modelling resulting from the application of our method can also contribute to detecting *schematic incorrectness* and *semantic misunderstandings* by means of automated reasoning. Ontology modellers, of SNOMED CT in this case, should pay special attention to false positives, because these are axioms proposed by our method that are not in SNOMED CT. With only three templates, our method suggests that 26 concepts (5%) in the congenital module and 62 concepts (15%) in the chronic module miss at least one axiom. These concepts should be analysed by SNOMED CT content editors to decide whether axioms are missing and if the ones suggested by our method are appropriate.

Our method is general, so it can be applied to other ontologies. We have performed a preliminary experiment with the Gene Ontology (GO) to analyse the potential reusability of the templates defined in this work. We have downloaded the February 2018 GO in OWL format, which contains 49 397 concepts, 107 736 logical axioms and 428 610 annotation assertions. GO is rich in LRs as a consequence of the use of naming conventions. We found 1277 LRs, of which 237 LRs (exhibited by 25 250 concepts) contain at least one preposition as token, and 144 LRs (exhibited by 16 714 concepts) of which are concepts in GO. Our templates 1 and 2 could be applied to the clusters whose LRs contain prepositions if the content around the preposition is a concept in the ontology, and template 3 could be applied to those whose LRs are concepts.

It should be noted that the domains of SNOMED CT and GO are different, so the property associated with, for example, the preposition 'of' may differ for both ontologies. This would mean that the template can be reused but the axiomatic pattern has to be adapted. One of the most frequent LRs in GO is related to regulation. The four most frequent LRs are: 'of' (13 928 concepts), 'regulation of' (10 966 concepts), 'positive regulation of' (3483 concepts) and 'negative regulation of' (3414 concepts). Moreover, GO provides modellers with eight types of properties (attributes in SNOMED CT terminology) including `regulates`, `positive regulates` and `negative regulates`. These could be used to adapt the axiomatic patterns associated with our template 2 shown in Table 8. A further analysis of the cluster alignments of these four LRs and their sub-regularities is needed. The analysis of how to use the other seven properties to define the templates for roughly 3000 concepts that exhibit 'of' but not the regulation would also be needed. This analysis could provide insights about the relation of the templates and the new relations created in the go-plus edition of GO [33].

4.3. Related work

In this paper we have presented a QA method based on the extraction of the semantics hidden in the content of the labels of concepts in biomedical ontologies and terminologies. Different approaches for the exploitation of hidden semantics can be found in the literature [8,10,12,34], including the application of the LSLD principle [14]. The analysis of term transformations is an ontology QA task in [8]. The definition of GO cross-products [12] exploited the compositional structure of GO concepts [10]. Later, unsupervised machine learning algorithms for combining lexical, syntactical and semantic regularities to define QA workflows for SNOMED CT were proposed [35]. Recently, inconsistencies in the logical definition of the concepts by contrasting lexical similarity and formal definitions in SNOMED CT were identified [34,36]. Missing hierarchical relations in SNOMED CT were identified from logical definitions based on the lexical features of concept names [37]. This method was recently expanded through the mining of non-lattice subgraphs [38]. The related work shows the importance of the lexical component for the QA of semantic resources, to which we want to contribute. Our approach can also be assimilated to the idea of exploiting a 'focus concept' and its neighbourhood presented in [17], but as novelty we propose the use of LRs as 'focus concepts'.

The enrichment of an ontology based on patterns was successfully applied with GO cross-products. GO cross-products were defined from the combination of automated tools and human curation for specified logical definitions for a large number of concepts. Our method innovates in how it automatically learns and extracts the lexical structure of labels in ontologies, obtains clusters of lexically similar labels and selects clusters based on statistical criteria. The alignment of the automatically selected clusters brings us closer to the automatic definition of the templates. This supposes an improvement with respect to the application of clustering techniques performed in [35], whose aim was to inform about unexpected situations. Similarly, the objective of the algorithms proposed in [34,36] is to detect inconsistencies in formal definitions of SNOMED CT concepts, but they cannot be considered pattern-based approaches. Other work [37,38] focused on detecting missing hierarchical relations based on lexical features of concepts. For example, they identify potential missing hierarchical relations: (1) 'Basal cell carcinoma of skin of lip' and 'Carcinoma of lip' or (2) 'Congenital vascular anomaly of eyelid' and 'Vascular anomaly of eyelid'. Our method would not be able to detect this first example, as LRs are defined considering consecutive tokens. However, the second example would be captured and solved using template 3 from Table 2.

4.4. Limitations and further work

The number of templates used in this paper has been limited by the need for their manual creation. This is the primary bottleneck of our method, which also makes the process not completely automatic, but has served to validate the applicability and usefulness of the method.

Despite this current limit of the applicability of the method on complete, large ontologies, this shortcoming can be overcome in the short term as indicated by the following evidence:

- Our clusters contain information for the automatic application of templates.
- Our clusters contain information useful for the definition of templates (see for instance the example in Section 2.3.1). We think that

it is possible to automatically extract the templates, so further research will primarily focus on this issue. We think that implementing aspects such as part-of-speech tagging on the labels would be helpful for this purpose.

- The templates defined in this work have been useful for the two modules extracted from SNOMED CT, which has to be interpreted in terms of reusability. The preliminary inspection of GO described in Section 4.2 also reinforces this idea. Therefore, the effort needed to create templates will decrease with the use of the method, especially if a library of axiomatic patterns is set and efficiently managed. We will also do research on their contextualisation as Ontology Design Patterns [39]. There are also two potential limitations concerning reusability: (1) as described in Section 4.2, some adaptations might be needed when reusing a template; (2) the capability of reusing templates in large ontologies collaboratively built by several experts may be reduced if the experts apply different modelling patterns.

We also think that using additional metrics for describing the LRs and the clusters such as the ones defined in [40] would help to select LRs which are potentially interesting and, therefore, would help to improve the results of the method.

We have developed another QA related method that we plan to combine with the present one to guide the developer in the actions to take to improve the ontology. In [41] we described a method for detecting potentially wrong axioms in ontologies by analysing whether the concepts that are lexically connected are also logically connected. If this method finds quality issues in a certain concept, the axiom proposed by the method presented in this paper should be helpful to improve the logical definition of such concept.

Finally, it would be necessary to develop a usable user interface to facilitate ontology developers to process the results of our methods and from which actions on the ontology could be implemented. We plan to include such an interface in OntoEnrich, through which this method will be made available to the community.

5. Conclusion

The large size and increasing importance of biomedical ontologies and terminologies such as SNOMED CT requires the development of QA methods that support the activity of content editors. We have presented a method for the extraction of axiomatic patterns from the content of the labels of concepts in biomedical ontologies and terminologies. The results are promising and shed light on how the lexical content can be used for the assurance of the quality of the ontology by following the LSLD principle. This work has benefited from the added value of techniques such as clustering. The semi-automatic method requires templates to be manually defined, so further work is needed to increase the degree of automation of the method.

Conflict of interest

The authors declared that there is no conflict of interest.

Acknowledgements

This work has been partially funded by the Spanish Ministry of Economy, Industry and Competitiveness, the European Regional Development Fund (ERDF) Programme and by the Fundación Séneca through grants TIN2014-53749-C2-2-R, TIN2017-85949-C2-1-R and 19371/PI/14. Philip van Damme was funded by the Erasmus +Traineeship Program.

References

[1] SemanticHealthNet, About the SemanticHealthNet project, 2017. http://www.semantichealthnet.eu.
[2] P.L. Whetzel, N.F. Noy, N.H. Shah, P.R. Alexander, C. Nyulas, T. Tudorache, M.A. Musen, 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 (2). http://dx.doi.org/10.1093/nar/gkr469.
[3] T.R. Gruber, A translation approach to portable ontology specifications, Knowl. Acquisit. 5 (2) (1993) 199–220.
[4] J. Geller, Y. Perl, M. Halper, R. Cornet, Special issue on auditing of terminologies, J. Biomed. Inform. 42 (3) (2009) 407–411, http://dx.doi.org/10.1016/j.jbi.2009.04.006 (auditing of Terminologies).
[5] X. Zhu, J.-W. Fan, D.M. Baorto, C. Weng, J.J. Cimino, A review of auditing methods applied to the content of controlled biomedical terminologies, J. Biomed. Inform. 42 (3) (2009) 413–425, http://dx.doi.org/10.1016/j.jbi.2009.03.003 (auditing of Terminologie).
[6] J.T. Fernández-Breis, M. Quesada-Martínez, A. Duque-Ramos, Can existing biomedical ontologies be more useful for EHR and CDS?, in: International Workshop on Knowledge Representation for Health Care, Springer, 2016, pp. 3–20.
[7] A. Third, Hidden semantics: what can we learn from the names in an ontology?, in: Proceedings of the Seventh International Natural Language Generation Conference, Association for Computational Linguistics, 2012, pp. 67–75.
[8] K. Verspoor, D. Dvorkin, K.B. Cohen, L. Hunter, Ontology quality assurance through analysis of term transformations, Bioinformatics 25 (12) (2009) i77–i84.
[9] C.M. Verspoor, C. Joslyn, G.J. Papcun, The gene ontology as a source of lexical semantic knowledge for a biological natural language processing application, in: SIGIR Workshop on Text Analysis and Search for Bioinformatics, 2003, pp. 51–56.
[10] P.V. Ogren, K.B. Cohen, G.K. Acquaah-Mensah, J. Eberlein, L. Hunter, The compositional structure of gene ontology terms, Biocomputing 2004, World Scientific, 2003, pp. 214–225.
[11] C.J. Mungall, Obol: integrating language and meaning in bio-ontologies, Compar. Funct. Genom. 5 (6–7) (2004) 509–520.
[12] C.J. Mungall, M. Bada, T.Z. Berardini, J. Deegan, A. Ireland, M.A. Harris, D.P. Hill, J. Lomax, Cross-product extensions of the gene ontology, J. Biomed. Inform. 44 (1) (2011) 80–86.
[13] A.L. Rector, R. Stevens, Quality assurance of the content of a large DL-based terminology using mixed lexical and semantic criteria: experience with SNOMED CT, in: Sixth Int. Conf. Knowl. capture, ACM, Banff, Alberta, Canada, 2011, pp. 57–64. http://dx.doi.org/10.1145/1999676.1999688.
[14] A.L. Rector, L. Iannone, Lexically suggest, logically define: quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT, J. Biomed. Inform. 45 (2) (2012) 199–209, http://dx.doi.org/10.1016/j.jbi.2011.10.002.
[15] M. Quesada-Martínez, J.T. Fernández-Breis, D. Karlsson, Suggesting missing relations in biomedical ontologies based on lexical regularities, Stud. Health Technol. Inform. 228 (2016) 384–388, http://dx.doi.org/10.3233/978-1-61499-678-1-384.
[16] M. Quesada-Martínez, J.T. Fernández-Breis, R. Stevens, N. Aussenac-Gilles, OntoEnrich: a platform for the lexical analysis of ontologies, in: G.C. Lambrix P., Blomqvist E., Qi G., Sattler U., Presutti V., Ding Y., Blomqvist E., Presutti V., Hyvonen E. (Ed.), Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 8982, Springer Verlag, Linkoping, 2015, pp. 172–176. http://dx.doi.org/10.1007/978-3-319-17966-7_25.
[17] C.P. Morrey, J. Geller, M. Halper, Y. Perl, The Neighborhood Auditing Tool: a hybrid interface for auditing the UMLS, J. Biomed. Inform. 42 (3) (2009) 468–489, http://dx.doi.org/10.1016/j.jbi.2009.01.006 (auditing of Terminologies).
[18] C. Manning, M. Surdeanu, J. Bauer, J. Finkel, S. Bethard, D. McClosky, The Stanford CoreNLP Natural Language Processing Toolkit, in: Proc. 52nd Annu. Meet. Assoc. Comput. Linguist. Syst. Demonstr, 2014, pp. 55–60. Available from: < arXiv:1011.1669v3 > . http://dx.doi.org/10.3115/v1/P14-5010.
[19] Universidad de Murcia, OntoEnrich web platform, 2017. http://sele.inf.um.es/ontoenrich.
[20] M. Horridge, S. Bechhofer, The OWL API: a Java API for OWL ontologies, Semant. Web 2 (1) (2011) 11–21, http://dx.doi.org/10.3233/SW-2011-0025.
[21] P. Jaccard, The distribution of the flora in the alpine zone, New Phytol. Trust 11 (2) (1912) 37–50.
[22] R. Suzuki, H. Shimodaira, Pvclust: an r package for assessing the uncertainty in hierarchical clustering, Bioinformatics 22 (12) (2006) 1540–1542.
[23] R. Suzuki, H. Shimodaira, pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap, 2015.
[24] D. Higgins, Multiple sequence alignment, Genetic Databases, Elsevier, 1997, pp. 165–183.
[25] A. Prlić, A. Yates, S.E. Bliven, P.W. Rose, J. Jacobsen, P.V. Troshin, M. Chapman, J. Gao, C.H. Koh, S. Foisy, R. Holland, G. Rimša, M.L. Heuer, H. Brandstätter-Müller, P.E. Bourne, S. Willis, BioJava: an open-source framework for bioinformatics in 2012, Bioinformatics 28 (20) (2012) 2693–2695, http://dx.doi.org/10.1093/bioinformatics/bts494.
[26] SNOMED International, SNOMED CT, 2017. https://snomed.org/snomed-ct.
[27] B.C. Grau, I. Horrocks, Y. Kazakov, U. Sattler, Modular reuse of ontologies: theory and practice, J. Artif. Int. Res. 31 (1) (2008) 273–318.
[28] E. Jiménez-Ruiz, B.C. Grau, U. Sattler, T. Schneider, R. Berlanga, Safe and Economic Re-Use of Ontologies: A Logic-Based Methodology and Tool Support, Springer, Berlin, Heidelberg, 2008, pp. 185–199, http://dx.doi.org/10.1007/978-3-540-68234-9_16.
[29] University of Manchester, Modularity: How do locality-based modules work? 2017. http://cs.owl.manchester.ac.uk/research/modularity.
[30] E. Mikroyannidi, R. Stevens, L. Iannone, A. Rector, Analysing syntactic regularities and irregularities in snomed-ct, J. Biomed. Semant. 3 (1) (2012) 8.
[31] P. López-García, S. Schulz, Can snomed ct be squeezed without losing its shape? J. Biomed. Semant. 7 (1) (2016) 56, http://dx.doi.org/10.1186/s13326-016-0101-1.
[32] A. Metke-Jimenez, M. Lawley, Snorocket 2.0: Concrete domains and concurrent classification, in: CEUR Workshop Proc., vol. 1015, 2013, pp. 32–38.
[33] The Gene Ontology Consortium, Expansion of the gene ontology knowledgebase and resources: the gene ontology consortium, Nucl. Acids Res. 45 (D1) (2017)

D331–D338, http://dx.doi.org/10.1093/nar/gkw1108.
[34] A. Agrawal, G. Elhanan, Contrasting lexical similarity and formal definitions in snomed ct: consistency and implications, J. Biomed. Inform. 47 (2014) 192–198.
[35] E. Mikroyannidi, M. Quesada-Martínez, D. Tsarkov, J.T. Fernández-Breis, R. Stevens, I. Palmisano, A quality assurance workflow for ontologies based on semantic regularities, in: 19th Int. Conf. Knowl. Eng. Knowl. Manag. EKAW 2014 8876, 2014, pp. 288–303. http://dx.doi.org/10.1007/978-3-319-13704-9_23.
[36] A. Agrawal, Y. Perl, C. Ochs, G. Elhanan, Algorithmic detection of inconsistent modeling among snomed ct concepts by combining lexical and structural indicators, in: 2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2015, pp. 476–483. http://dx.doi.org/10.1109/BIBM.2015.7359731.
[37] O. Bodenreider, Identifying missing hierarchical relations in snomed ct from logical definitions based on the lexical features of concept names, in: International Conference on Biomedical Ontology and BioCreative (ICBO BioCreative 2016), Proceedings of the Joint International Conference on Biological Ontology and BioCreative (2016), CEUR-ws.org Volume 1747, CEUR-ws.org Volume 1747, 2016.
[38] L. Cui, O. Bodenreider, J. Shi, G.-Q. Zhang, Auditing SNOMED CT hierarchical relations based on lexical features of concepts in non-lattice subgraphs, J. Biomed. Inform. 78 (October 2017) (2018) 177–184, http://dx.doi.org/10.1016/j.jbi.2017.12.010.
[39] Ontology Design Patterns, 2018. http://ontologydesignpatterns.org.
[40] M. Quesada-Martínez, E. Mikroyannidi, J.T. Fernández-Breis, R. Stevens, Approaching the axiomatic enrichment of the Gene Ontology from a lexical perspective, Artif. Intell. Med. 65 (1) (2015) 35–48, http://dx.doi.org/10.1016/j.artmed.2014.09.003.
[41] M. Quesada-Martinez, J.T. Fernandez-Breis, D. Karlsson, Suggesting missing relations in biomedical ontologies based on lexical regularities, Stud. Health Technol. Inform. 228 (2016) 384–388, http://dx.doi.org/10.3233/978-1-61499-678-1-384.