

# Generation of Heterogeneous Semantic Annotations of XML Pages: a Multi-viewpoints Approach

Meriem Djezzar

University of Abbes Laghrour, Khenchela, Algeria

LIRE Laboratory, Constantine 2 - Abdelhamid Mehri -University, Constantine, Algeria

E-mail: meriem.djezzar@univ-khenchela.dz

Mounir Hemam

Khenchela University, ICOSI Laboratory, BP 1252 El Houria, 40004 Khenchela, Algeria

E-mail: hemam.mounir@univ-khenchela.dz

Djama Ouahiba

Lire Laboratory, University of Abdelhamid Mehri Constantine 2, Ali Mendjeli, Constantine, Algeria

E-mail: djama\_ouahiba@umc.edu.dz

Zianou Ahmed Seghir

Khenchela University, ICOSI Laboratory, BP 1252 El Houria, 40004 Khenchela, Algeria

E-mail: zianou\_ahmed\_seghir@yahoo.fr

**Keywords:** XML page, semantic web, semantic annotation, multi-viewpoints ontology, consensual semantic annotation, heterogeneous semantic annotations

**Received:** August 1, 2021

*The semantic annotations presented in current research are based on consensual descriptions of domain knowledge which are used to generate a consensual interpretation of the resource content. Sometimes specialists in a domain want to enrich this interpretation with specific interpretations based on their specialties, consistent with the interpretations of other specialists. However, extant work cannot enrich consensual annotations with heterogeneous annotations. We propose an approach that constructs a consensual (global) annotation and then enriches the consensual annotation with heterogeneous (local) annotations, each of which is a specification of the global annotation. The heterogeneous annotations are consistent with the global interpretations and with one another. What is unique in our approach is that one element in a resource can be annotated with different specific concepts, based on different viewpoints. Our approach utilizes a multi-viewpoints ontology [1], with heterogeneity at the local level and consensus at the global level.*

*Povzetek: Predstavljena je metoda za generiranje heterogenih semantičnih anotacij XML strani na osnovi različnih, a skladnih pogledov*

## 1 Introduction

To equip resources with explicit meanings which can be interpreted by software, it is necessary to associate these resources with semantic annotations [2]. In the semantic web, semantic annotation is defined as ‘an approach to link ontologies to the original information sources’ [3]. An ontology provides consensual descriptions of the elements of a domain [4, 5]. Semantic annotations are based on these consensual descriptions to generate a consensual interpretation of the resource content.

Generally, in any given domain, there are different specialists. All of these specialists share a common description of the domain knowledge (a consensual description). In addition, each specialist has a specific description according to his or her specialty (his or her local viewpoint).

Current work such as [6, 7, 8, 9, 10, 11, 46] constructs only a consensual annotation of a resource, using a single ontology. Other work that constructs only a consensual semantic annotation of a resource, such as [12, 13, 14, 15], uses multiple ontologies, intending to obtain a ‘bigger’ ontology with a larger set of concepts. Trichet, et al. [16] have developed a platform that allows users to utilize one or more ontologies to generate their own annotations for the same resource.

However, this approach allows constructing only heterogeneous annotations without a common, consensual annotation. In contrast to these works, we are interested in an annotation system with the following characteristics:

- it should include a consensual (global) annotation;
- the consensual annotation can be enriched by many specific local annotations (heterogeneous annotations) which are mutually consistent;

- the local annotations can be enriched while maintaining consistency between them; and
- a local semantic annotation of a page's content is generated for each different viewpoint.
- However, we cannot accomplish these goals using a classical ontology (a mono-viewpoint ontology), because classical ontologies can generate only consensual annotations. Moreover, multi-ontologies allow only the construction of heterogeneous annotations, without a consensual annotation, so that it is necessary to reason about the alignment of disjoint ontologies.

So, to accomplish our goals, we exploit the multi-viewpoint ontology developed in [1]. A multi-viewpoints ontology is cohabited by heterogeneity at the local level and consensus at the global level. The knowledge (concepts and roles) common to all viewpoints is described in the global representation. The local knowledge for each viewpoint is described in a local representation, and the local representations are linked by bridge rules.

In this work, we search for the page elements that can be annotated according to the global viewpoint, in order to generate a global (consensual) annotation. Next, for each viewpoint, we search the page for the local attributes of the global concepts in order to identify the local concepts in the page that specify (are subsumed by) the global concepts. In addition, we will focus on the annotation of the DOMXML tree using a multi-viewpoints ontology. Each node in the DOMXML tree can be annotated (labeled) by an ontological element. We build a labeled DOMXML subtree for each viewpoint that contains the relevant page elements as seen from this viewpoint. Thus, a single node can be annotated by different specific concepts according to different viewpoints.

The remainder of this paper is organized as follows. Section 2 gives some terminology and definitions. Then, in section 3, we show how to link each DOMXML node to the appropriate ontological element. In section 4, we explain our proposed approach. In section 5, we present the formalization and implementation. Section 6 presents a discussion, and section 7 concludes this paper and talks about further work.

## 2 Terminology and definitions

### 2.1 The viewpoint approach: an overview

For a given domain of knowledge, several criteria can be used to observe an object. These different perceptions of the world are called viewpoints or perspectives. In computer science, most of data modeling systems don't deal with the variety of perceptions related to the same universe of discourse and develop tools to create a single model for a single vision of the observed world. The viewpoint approach is opposed to this monolithic approach and makes it possible to model the same reality according to different points of view.

Several interpretations of viewpoint notion are possible. One of the first references to viewpoints was

proposed by Minsky [17]: viewpoints correspond to different perceptions of an object with respect to observer's position. The second interpretation is a knowledge domain one: viewpoints correspond to the different ways to translate knowledge with respect to the social position, know-how and competence of an expert. In this interpretation, a viewpoint includes context and the perception of a person or group of persons [18].

The viewpoint mechanism has been integrated into various contexts and used to solve different problems. In the following, we identify the main objectives in integrating viewpoints into computer systems. Note that there is no single use of this concept that includes all of these objectives.

- The viewpoint as a means of providing multiple descriptions of an entity: the viewpoint concept seems to naturally result from the multiple views of objects of a specific study. As a matter of fact, a real-world entity can have many behavioral contexts and many states from which the notion of multiple descriptions has been derived. In this case, it is defined as the fact of conferring several partial descriptions to the same universe of discourse each of which describes it in a given viewpoint.
- The viewpoint as a means of mastering system complexity: several research works are based on the viewpoint concept with the principal objective of explicitly taking into account the complexity of the system. The result of the study is then held by dividing it into partial descriptions according to different and complementary aspects.
- The viewpoint as an approach for the modeling and distributed development of systems: many authors state that the modelling of complex systems cannot be handled with the same techniques as used for simple systems. Different works suggest a distributed development approach based on viewpoint notion. Hence, every development process can be represented by correlated viewpoints.

### 2.2 Multi-viewpoints ontology

Hemam et al. [18] and Hemam [1] investigated the problem of representing a domain ontology which takes into account the notion of a viewpoint. This type of ontology, which they call a Multi-Viewpoints Ontology (MvpOnt), is a multiple description of a single universe of discourse according to various viewpoints.

The ontology is defined as a 4-tuple  $O = (CG, RG, VP, M)$ , where CG is a set of global concepts, RG is a set of global roles, VP is a set of viewpoints and M is a set of bridge rules.

A viewpoint, which is a partial description of a universe of discourse from a particular perspective, is defined as a triple  $VPK = (CL, RL, AL)$ , where CL is a set of local concepts, RL is a set of local roles and AL is a set of local individuals. Some fundamental definitions follow:

- A global concept models a generic family in the real world and groups a set of individuals. Each global concept can be described according to different viewpoints.
- A local concept is a concept that is seen and described locally, according to a particular viewpoint.
- A global role is a relationship between two local concepts that are defined in two different perspectives.
- A local role is a relationship between two local concepts defined in the same viewpoint.
- A bridge rule represents consensual relationships between two local concepts or two local roles represented in two different viewpoints.

Under a particular viewpoint, each particular individual is attached to only one local concept. This is motivated by the argument that an individual belonging to two unrelated local concepts can only do so under different viewpoints [45]. This organization is very useful since, from a particular viewpoint, one can classify an individual with regard to only the relevant properties [43]. As a consequence, the set of concepts to be considered will be small, whereas considering the entire lattice of possible structures would confuse the user.

### 3 Linking DOMXML nodes to ontological elements

Apparo and Pixley [19] noted that DOM is a specification of the W3C (World Wide Web Consortium) that allows defining the hierarchical structure of an XML document by a tree in which:

- each XML tag is represented by an element node with the same name as the XML tag,
- the imbrications between tags are represented by edges,
- the value of an XML tag is represented by a text node, and
- the attribute of an XML tag is represented by an attribute node.

DOM has several methods for manipulating XML files, such as:

- RemoveChild, which allows deleting a target node, and
- hasChildNodes, which returns true if the node has children, and false otherwise.

In the following figure, we present an example of an XML representation. This representation shows a set of apartments, each of which is defined by tenant, location and rent:

A DOMXML tree is composed of several levels. The first level contains the element nodes which are parents of

```
<Apartments>
  <Apartments Adress= "bel air street Constantine city, 25000" >
    <Tenant> Meriam </Tenant>
    <Location> Downtown </Location>
    <Rent> 500.00 </Rent>
  </Apartments>
```

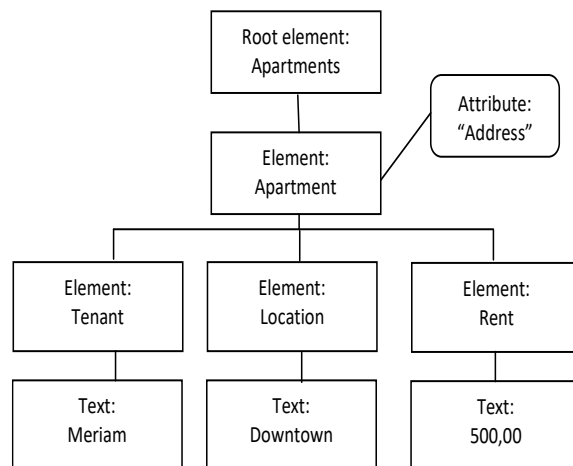


Figure 1: Example of a graphical DOMXML representation.

text nodes. The second level contains the element nodes which are parents of at least one element node at the first level. And so on: level  $L_i$  contains the element nodes which are parents of at least one element node at level  $L_{i-1}$ .

In the following, we will describe how to annotate each node with an element defined in the multi-viewpoints ontology. First, however, we show how to link a node of a DOMXML tree to the appropriate ontological element, which is represented by the OWL language.

We compare the node in DOMXML with the OWL representation of the ontological element at the structural and terminological levels. At the terminological level, we compare the name of the node with the name of the ontological element (as a string comparison). We use one of the metrics which are defined in the literature [20, 21, 22, 23]. If the names are technical words, we search for an identity between the names or between their synonyms. We use technical dictionaries, such as UMLS [24] for the biomedical domain, to search for synonyms of technical words. If the names aren't technical words, we search for a similarity by searching for synonyms of the ontological element in WordNet [25].

At the structural level, Table 1 compares the DOMXML structure of the node with the OWL structure of the ontological element. There are several approaches to mapping an XML structure to OWL. Ferdinand et al. [26] and Yahia et al. [27] generate an ontology from XML based on mapping rules. Bohring and Auer [44] proposed mapping rules to map XML tags to OWL. These mapping rules show how to link a node to the appropriate ontological element, according to the structure of this node. In the DOMXML tree, each element node can have one of the following structures:

### 4 Presentation of the proposed approach

The annotation of the DOMXML tree is done by exploiting a multi-viewpoint ontology and consists of the following three phases.

Case	Element node	Converted to	Annotated by
1	Element node parent of text node which is not attached to an attribute node	<i>owl: DatatypeProperty</i>	Instance of an ontological attribute of the concept.
2	Element node parent of text node attached to at least one attribute node	<i>owl: Class</i>	Ontological concept defined by a set of attributes.
3	Element node without text node and which is not attached to an attribute node	<i>owl: Class</i>	Primitive concept defined only by its name.
4	Element node without text node which is attached to at least one attribute node	<i>owl: Class</i> ; its attribute nodes are converted to <i>owl: DatatypeProperty</i>	Ontological concept defined by a set of attributes.
5	Element node with children Nodes	<i>owl: Class</i>	Concept defined by other concepts. We search for the concept such that the ontological elements defining this concept annotate the children of this element node.

Table 1: Different structures of DOMXML nodes.

#### 4.1 Phase 1: Construction of the original DOMXML tree

The first phase consists of building the DOMXML tree for the resource page. We make  $n$  copies of this DOMXML tree, where  $n$  is the number of the viewpoints seen in the domain, plus one copy for the global annotation.

#### 4.2 Phase 2: Generation of the global annotation

In this phase, we will build the annotated DOMXML subtree from the copy of the original DOMXML tree. We do this in two steps:

##### 4.2.1 Step 1: Generation of the annotation

In this step, we annotate each node with a global ontological element, using the method described in the previous section.

The unique characteristic of the multi-viewpoints ontology is that a global concept is a concept which is defined by a set of global attributes and a set of local attributes, and a subset of the global attributes comprises the key attributes for this concept. So the annotation of an element node by a global concept is related to the annotation of some nodes which are attached to this element node (its text node, its attributes node and its children) by the key attributes of this concept.

##### 4.2.2 Step 2: Deletion of irrelevant nodes

A node is relevant to the global viewpoint if we can find its annotation; otherwise the node is irrelevant.

Irrelevant nodes are removed if they are attribute nodes, text nodes or element nodes which do not have annotated children or annotated text nodes.

Irrelevant nodes which are element nodes with annotated children or annotated text nodes are kept in order to retain relevant nodes, because there are relationships between these children in the page. However, there are no roles between the ontological elements which annotate these children.

At the end of this phase, we have an annotated (labeled) DOMXML subtree in accordance with the global viewpoint.

#### 4.3 Phase 3: Generation of the local annotations

In this phase, we will build a subtree  $DOMVP_i$  from a copy of the original DOMXML tree and the labeled subtree with the global viewpoint for each viewpoint  $VP_i$ .  $DOMVP_i$  will be a labelled DOMXML subtree with the viewpoint  $VP_i$ . Initially, each  $DOMVP_i$  is a copy of the original DOMXML tree.

We generate the annotations level by level in each subtree  $DOMVP_i$ . We begin with  $L_j$  as the first level and then move up one level at a time until we reach the root. This procedure involves the following steps:

##### 4.3.1 Step 1: Generation of the annotations

Some nodes are already annotated in the global viewpoint, and we save the global annotations of these nodes. We annotate the other nodes with local ontological attributes, according to viewpoint  $VP_i$ , by using the method described in the Table 1.

Then, for each node  $N_k$  annotated by a global concept  $G_k$ , we see whether we can find a local concept specifying (subsumed by) this global concept which can be used to annotate  $N_k$  according to  $VP_i$ . Note that the node  $N_k$  has some children which have been annotated (in phase 2) by global attributes (and other children which have been annotated in phase 3 with local attributes according to  $VP_i$ ). The children which are annotated by local attributes help in finding the local concepts which specify the concept  $G_k$ .

If we can't find a specific local concept, the global annotation of  $N_k$  will not be changed. Thus, each node at level  $L_j$  of the subtree  $DOMVP_i$  is annotated either by a global ontological element or by a local ontological element according to  $VP_i$ .

### 4.3.2 Step 2: Propagation of the annotations

In the multi-viewpoints ontology, the local knowledge of each viewpoint is described in the local representation. The local representations are linked by bridge rules. Bridge rules allow communication to be established between the local presentations of the local knowledge. The bridge rules can link two concepts of different viewpoints. This allows new local individuals (instances of concepts) to be inferred in different viewpoints as well as detection of contradictions at the level of individuals.

Annotation of an element on a resource page with an ontological concept means the creation of an individual instance of this concept. In this section, we explain how to apply the bridge rules to the annotations obtained at level  $L_j$  of each subtree  $DOMVP_i$ . This allows local annotations to be propagated between all the subtrees at level  $L_j$ . Propagation of the annotations will allow the inference of new local annotations and detection of contradictions in all the subtrees at level  $L_j$ . For this, we generate a rule-based system composed of a knowledge base. The latter is composed of facts and production rules (inference rules):

**Construction of the facts.** The local annotations which have been obtained allow creating instances of some ontological concepts. These local annotations will be transformed to facts.

As an example, if the node NE is annotated by the concept C according to  $VP_i$ , we can create an instance of C from NE as follows:

`is_instance (NE,  $VP_i$ : C)`

**Construction of the production rules.** The production rules will be built from the bridge rules which are defined in the multi-viewpoints ontology. A bridge rule is a statement of one of the three following forms:

- An inclusion bridge rule, noted  $X:vp_i \rightarrow Y:vp_j$ , means that an individual which is an instance of the source concept X under the  $vp_i$  is also an instance of the target concept Y under the  $vp_j$ .
- A bi-directional inclusion bridge rule, noted  $X:vp_i \leftrightarrow Y:vp_j$ , means that the sets of possible extensions of the two local concepts under different viewpoints are equal.
- A bi-directional exclusion bridge rule, noted  $X:vp_i \rightarrow \perp Y:vp_j$ , means that the two local concepts X and Y cannot be at the same time representations of the same individual.

Each of the previous types of bridge rules will be transformed to a production rule as follows:

- A inclusion bridge rule will be transformed to a production rule that has:  
Premise: X is an individual instance of the concept C1 according to the viewpoint  $VP_1$ , and  
Conclusion: X is an individual instance of the concept C2 according to the viewpoint  $VP_2$ .  
`if is_instance (X,  $VP_1$ :C1) then`  
`is_instance (X,  $VP_1$ :C2)`
- A bi-directional inclusion bridge rule will be transformed to two production rules. The first rule has  
Premise: X is an individual instance of the

concept C1 according to the viewpoint  $VP_1$ , and  
Conclusion: X is an individual instance of the concept C2 according to the viewpoint  $VP_2$ .

`if is_instance (X,  $VP_1$ :C1) then`  
`is_instance (X,  $VP_2$ :C2)`

- The second rule has:  
Premise: X is an individual instance of the concept C2 according to the viewpoint  $VP_1$ , and  
Conclusion: X is an individual instance of the concept C1 according to the viewpoint  $VP_2$ .  
`if is_instance (X,  $VP_2$ :C2) then`  
`is_instance (X,  $VP_1$ :C1)`
- A bidirectional exclusion bridge rule will be transformed to a production rule that has:  
Premise: X is an individual instance of the concept C1 according to the viewpoint  $VP_1$ , and  
Conclusion: X is not an individual instance of the concept C2 according to the viewpoint  $VP_2$ .  
`if is_instance (X,  $VP_2$ :C2) then`  
`¬is_instance (X,  $VP_1$ :C1)`

## 5 Formalisation and implementation

The implementation of the proposed approach consists of generating a system that is composed of a set of modules. Each module allows implementing a phase or a step of the proposed approach.

In order to represent the multi-viewpoints ontology, we have used the MVP-OWL, proposed in [28] as an extension of OWL. So, the multi-viewpoints ontology is saved in an MVP-OWL file. Both the owl file and the MVP-OWL file have an XML/RDF structure. So, these files can be manipulated as XML files or RDF files. The DOMXML model (Document Object Model) allows accessing and manipulating the contents of the XML file with PHP language. So, the XML structure of the MVP-OWL file can be exploited by the DOMXML model in PHP language. The obtained multi-viewpoints annotation will be saved in RDF files.

The new constructs, used to represent multi-viewpoints ontological knowledge, are described in the following:

- `vpowl:Viewpoint` that is used to represent viewpoints.
- `vpowl:GlobalClass` that is used to represent global concepts.
- `vpowl:LocalClass` that is used to represent local concepts.
- `vpowl:GlobalProperty` that is used to represent global roles.
- `vpowl:LocalProperty` that is used to represent local roles.
- `vpowl:underViewpoint` that is used to specify the viewpoint of the local concept.
- `vpowl:onViewpoint` that is used to specify the viewpoint of the local role and local attribute.
- `vpowl:belongtoViewpoint` that is used to specify the viewpoint of the local individual.
- `vpowl:InclusionBridge`, `vpowl:EquivalenceBridge` and `vpowl:ExclusionBridge` are used to represent bridge rules.

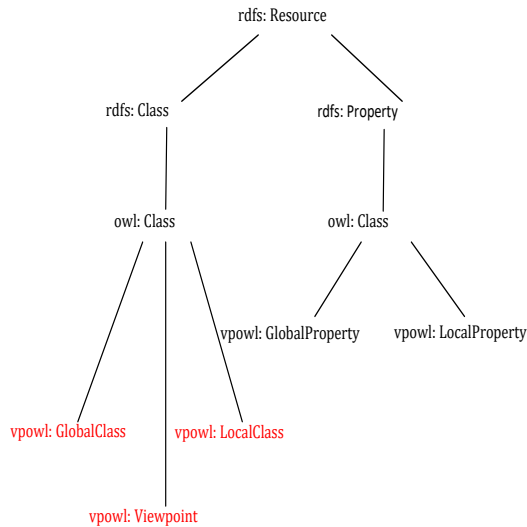


Figure 2: Subclass relationships between mvpowl and owl.

Production rules	Formalisation by SWRL rules
<i>if is_instance (X, VP1:C1) then is_instance (X, VP1:C2)</i>	<i>VP1:C1(?X) → VP1:C2(?X)</i>
<i>if is_instance (X, VP2:C2) then ¬is_instance (X, VP1:C1)</i>	<i>VP1:C1(?X) → (notVP1:C2)(?X)</i>

Table 2: Formalisation of some production rules by SWRL rules.

These constructs are represented as subclasses of OWL classes. For example, *vpowl:GlobalClass* is represented as a subclass of *owl:Class*, as shown in Fig 2.

In addition, the production rules can be formalized as SWRL rules [43]. These rules are expressed in terms of the multi-viewpoint ontology concepts (i.e. classes, properties, individuals) and description logic, using a stamping technique. Table 2 shows some examples.

## 6 Results and discussion

This section evaluates our proposed process, shows the place of our work in the literary and compares our work with the existing works.

Several approaches to semantic annotation of text have been developed, for example, Baumgartner et al. [6], Carme et al. [29], Ciravegna and Wilks [30], Davulcu et al. [31], Euzenat [32], Hanndschuh et al. [7], Petridis et al. [33], Rahayu [8], Sitthisarn and Bahoh [9] and Uren et al. [10]. These approaches are based on natural language processing to extract data categories. However, it is difficult to identify data categories because of ambiguities.

Semi-structured documents represent the hierarchical structure of data categories. Several approaches in this direction have been developed, for example, Alani et al. [34], Hignette et al. [35], Thiam [36] and Zhang et al. [11]. These approaches are based on exploiting the structure of the resource page. But they are also based on classical

ontology and cannot take different viewpoints into consideration.

The work on multi-viewpoints is divided between two directions. The first direction represents multiple viewpoints by using disjoint ontologies. Each viewpoint is represented by an ontology, and the ontologies are linked by bridge rules or context mappings to establish interconnections between them. These authors are interested in the mechanisms which allow linking available ontologies. The work in this direction includes Borgida and Serafini [37], the model C-language OWL [38], and the MPV model [39]. This approach requires the alignment and matching of ontologies.

Some work has proposed semantic annotation based on multi-ontologies. The aim of using the multi-ontologies is to obtain a bigger ontology that contains a large set of concepts, this work includes, for example, Bhogal et al. [40], Gómez- Berbís et al. [13], Mena et al. [14] and Wang et al. [15]. These authors use one ontology to annotate some page elements, and they use another ontology to annotate other page elements because the first one is insufficient. Thus, the aim is not to generate a semantic annotation for each viewpoint, but rather to generate one semantic annotation that groups together all the knowledge of the different domains. In contrast, in our work, we aim to generate an annotation for each viewpoint with its own interpretation of the resource content based on its own knowledge.

Trichet, Aimé and Thovex [16] presented a platform that allows the generation of semantic annotations by taking multiple viewpoints and multiple users into consideration. The different viewpoints are defined through the use of multiple ontologies. These ontologies can be related to connected or unconnected domains. Users can use one or more ontologies to generate their annotations according to their own angle. In this work, one local annotation is manually generated for each user (viewpoint). This approach allows generating only heterogeneous annotations, without a consensual annotation.

In our previous work [42], we proposed an approach to multi-viewpoints semantic annotation based on annotation classes. An annotation class aims to link a page element to an ontological element according to a given viewpoint. The annotation class is manually generated by experts in the domain. The multi-viewpoints semantic annotation is the multi-instantiation of these classes. This approach requires unifying the structures of resource pages in the same domain. The modification of the structures is based on modifying the annotation processes.

Our current work differs from our earlier work in exploiting all the notions (global concepts, local concepts, subsumption between global and local concepts, ...) which are defined in the multi-viewpoints ontology, in order to generate the consensual annotation and the heterogeneous annotations that specify the consensual annotation (i.e., the generation of each local annotation is not independent of the global annotation).

Additionally, our current work differs from our previous work in its use of the bridge rules that are defined in the multi-viewpoints ontology to develop a set of rules

that allow propagating the local annotations. This propagation enables enrichment of the local annotations, and it also enables establishing consistency (i.e. correcting contradictions) between the local annotations.

It should be noted that, the execution time optimization is an advancement of our approach compared to the consummation of execution time in the case of the exploitation of a classical ontology. In fact, the local annotations according to different involved viewpoints can be realized in parallel (simultaneously).

According to Figure 3, most of the time is consumed during the local annotations phase because generally, the local hierarchies of different viewpoints are longer than the global hierarchy in the multi-view ontology. In addition, the time of the propagation of annotations, by using bridge rules, is negligible compared with local annotations phase. On the other hand, the time of the global annotation phase is very negligible compared with the second and third phases.

To evaluate our current approach, we applied it in several domains, including real estate agency, library, and medical domains. We had experts in the domains manually generate multi-viewpoints semantic

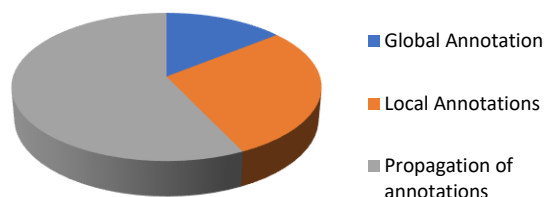


Figure 3: Distribution of time consumption between global annotation, local annotations and propagation of local annotations steps.

annotations, and we compared the results of our approach with the experts’ results, calculating the error rate in order to evaluate our results. The error rate is the number of erroneous annotations (compared to the experts’ results) divided by the total number of annotations. We summarise our results in Table 3:

It can be seen that the average error rates are very low. We therefore conclude that we have obtained good results with our approach.

Domain	Ontology size (Number of concepts)	Number of annotated pages	Number of viewpoints	The error rate for the global annotation	The error rate for the local annotation
Real estate	30	200	3	14%	19.05%
Library	50	100	5	23.48%	24.32%
Medical	25186	2000	7	27.5%	29.12%

Table 3: Experimental Results.

Domain	Ontology size (Number of concepts)	Number of annotated pages	Time consumed with classical ontology	Time consumed with multi-viewpoints ontology (phase 1 + phase 2 + phase 3)	Time consumed with multi-viewpoints ontology (phase 1)	Time consumed with multi-viewpoints ontology (phase 2)	Time consumed with multi-viewpoints ontology (phase 3)
Real estate	30	200	5,73 second	1.75 second	0,31 second	0.51 second	0.92 second
Library	150	100	19,25 second	5.67 second	1.20 second	1.6 second	2.87 second
Medical	25186	2000	850 second	774.41 second	205.2 second	251.86 second	317.35 second

Table 4: The average consumption of execution time to annotate the resource elements with ontology elements (time is measured in seconds).

Domain	Ontology size (Number of concepts)	Number of annotated pages	Number of viewpoints	Number of enriched global annotations	Number of enriched local annotations
Real estate	30	200	3	5,73 second	1.75 second
Library	150	100	5	19,25 second	5.67 second
Medical	25186	2000	7	850 second	774.41 second

Table 5: Evaluation of the capacity of the proposed approach in the enrichment of annotations

In Table 4, we present a comparison, of the average consumption of execution time, between our approach and the existing works that allow exploiting a classical ontology.

We also calculated the number of specific annotations that enriched the consensual annotation and the number of inferred local annotations in order to show the advantages of our approach with respect to its inference capacity. Table 5 summarises the results.

## 7 Conclusion

We have proposed an approach to multi-viewpoints semantic annotation based on the DOMXML structure of a resource page and the multi-viewpoints ontology. This approach allows:

1. generating a consensual (global and common) annotation,
2. enriching the local annotations via an inference mechanism using a rule-based system,
3. generating a local annotation for each viewpoint via labeled substructures,
4. minimizing the number of comparisons between the page elements and the ontological elements by exploiting the structure of each page element and the rules mapping XML to OWL,
5. letting one page element be annotated by different specific concepts according to different viewpoints, and
6. letting a page element that doesn't have a local annotation according to a given viewpoint keep its global annotation.

For future work, we plan to introduce the notion of fuzziness in the search for the attributes of concepts in the page. We also propose developing an approach that allows the enrichment of a multi-viewpoints ontology by multi-viewpoints semantic annotation of resources, as well as an approach to reasoning on the multi-viewpoints annotations which allows selecting the appropriate results in a search engine according to a particular user viewpoint. Finally, we hope that our work on multi-viewpoints annotation can foster ideas for future investigations.

## Acknowledgement

We would like to thank the anonymous referees for their valuable comments on an earlier version of this paper.

## References

- [1] Hemam, M. (2018). An extension of the ontology web language with multi-viewpoints and probabilistic reasoning. *Int. Journal of Advanced Intelligence Paradigms*, Inderscience Publishers Ltd. Vol. 10, N° 3, pp. 247–265. <https://doi.org/10.1504/ijaip.2018.090789>
- [2] Bechhofer, S., Carr, L., Goble, G., Kampa, S., Miles, B.T. (2002). The semantics of semantic annotation, in Meersman, R. and Tari, Z. (Eds.), *On the Move to Meaningful Internet Systems 2002: CoopIS, DOA, and ODBASE*, Springer-Verlag, Berlin, pp. 1152–1167. [https://doi.org/10.1007/3-540-36124-3\\_73](https://doi.org/10.1007/3-540-36124-3_73)
- [3] Lin, Y., Krogstie, J. (2008). Semantic annotation for process models: Facilitating process knowledge management, *International Journal of Information System Modelling and Design*, Vol. 1 No. 3, pp. 45–67. <https://doi.org/10.4018/jismd.2010070103>
- [4] Abecker, A. and Van-Elst, L. (2004). Ontologies for knowledge management, in Staab, S. and Studer, R. (Eds.), *Handbook on Ontologies*, Berlin: Springer-Verlag, pp. 435–454. [https://doi.org/10.1007/978-3-540-24750-0\\_22](https://doi.org/10.1007/978-3-540-24750-0_22)
- [5] Gruber, T.R. (1995). Towards principles for the design of ontologies used for knowledge sharing, *International Journal of Human and Computer Studies*, Vol. 43 No. 5, pp. 907–928. <https://doi.org/10.1006/ijhc.1995.1081>
- [6] Baumgartner, R., Flesca, S., Gottlob, G., (2001). Visual web information extraction with lixto, *Very Large Data Base Endowment Inc*, Vol. 10 No. 1, pp. 119–128.
- [7] Handschuh, S., Staab, S. and Ciravegna, F. (2002). S-cream-semi-automatic creation of metadata, in Gómez-Pérez, A. and Benjamins, V.R. (Eds.), *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, Springer-Verlag, Berlin, pp. 358–372. [https://doi.org/10.1007/3-540-45810-7\\_32](https://doi.org/10.1007/3-540-45810-7_32)
- [8] Rahayu, S.B., (2012). Ranking algorithm for semantic document annotations, *International Journal of Information Retrieval Research*, Vol. 2 No. 1, pp. 1–10. <https://doi.org/10.4018/ijirr.2012010101>
- [9] Sitthisarn S., Bahoh B. (2014). Towards Automatic Semantic Annotation of Thai Official Correspondence: Leave of Absence Case Study. In: Boonkrong S., Unger H., Meesad P. (eds) *Recent Advances in Information and Communication Technology. Advances in Intelligent Systems and Computing*, vol 265. Springer, Cham. [https://doi.org/10.1007/978-3-319-06538-0\\_27](https://doi.org/10.1007/978-3-319-06538-0_27)
- [10] Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., & Ciravegna, F. (2006). Semantic annotation for knowledge management: requirements and a survey of the state of the art. *Journal of Web Semantics*, 4(1), 14–28. <https://doi.org/10.1016/j.websem.2005.10.002>
- [11] Zhang, K., Li, J., Hong, M., Yan, X., Song, Q. (2014). Semantics enabled intelligent semi-structured document processor, in Yuan, Y., Wu, X. and Lu, Y. (Eds.), *Trustworthy Computing and Services*, Springer-Verlag, Berlin, pp. 328–344. [https://doi.org/10.1007/978-3-662-43908-1\\_41](https://doi.org/10.1007/978-3-662-43908-1_41)
- [12] Bhogal, J., Macfarlane, A. and Smith, P. (2007). 'A review of ontology based query expansion', *Information Processing & Management*, Vol. 43 No. 4, pp. 866–886. <https://doi.org/10.1016/j.ipm.2006.09.003>
- [13] Gómez-Berbís, J.M., Colomo-Palacios, R., López-Cua drado, J.L., González-Carrasco, I. and García-Crespo, Á. (2011). 'SEAN: Multi-ontology sema



- ntic annotation for highly accurate closed domains', *International Journal of the Physical Sciences*, Vol. 6 No. 6, pp. 1440–1451.
- [14] Mena, E., Kashyap, V., Illarramendi, A., Sheth, A. (2000). Imprecise answers in distributed environments: Estimation of information loss for multi-ontology based query processing, *International Journal of Cooperative Information Systems*, Vol. 9 No. 4, pp. 403–425.  
<https://doi.org/10.1142/s0218843000000193>
- [15] Wang, P., Xu, B., Lu, J., Li, Y., Jiang, J. (2004). Bridge ontology: A multi ontologies-based approach for semantic annotation, *Wuhan University Journal of Natural Sciences*, Vol. 9 No. 5, pp. 617–622.  
<https://doi.org/10.1007/bf02831652>
- [16] Trichet, F., Aimé, X. and Thovex, C. (2010). OSIRIS: Ontology-based system for semantic information retrieval and indexation dedicated to community and open web spaces, in Blanchard, E.G. and Allard, D. (Eds.), *Handbook of Research in Culturally-Aware Information Technology: Perspectives and Models*, Hershey, PA, pp. 465–483.  
<https://doi.org/10.4018/978-1-61520-883-8.ch021>
- [17] Minsky, M., (1975) A framework for representing knowledge. In Winston, P., ed.: *The Psychology of Computer Vision*. McGraw Hill, New York.
- [18] Hemam, M., Djezzar M. and Boufaïda, Z. (2017). Multi-viewpoint Ontological Representation of Composite Concepts : a Description Logics Based Approach. *Int. Journal of Intelligent Information and Database Systems*, Inderscience Publishers Ltd. Vol. 10, N° 1/2, pp. 51–68.  
<https://doi.org/10.1504/ijjids.2017.086193>
- [19] Apparo, V. and Pixley, T. (2008). 'The NGLayout Document Object Model (DOM)' Roadmap. Mozilla Organization.  
<http://www.mozilla.org/newlayout/dom-roadmap.html>.
- [20] Camacho, H. and Salhi, A. (2006). A string metric based on a one-to-one greedy matching algorithm, *Research in Computing Science*, Vol. 19, pp. 171–182.
- [21] Hadjieleftheriou, M. and Srivastava, D. (2010). Weighted set-based string similarity, *Data Engineering Bulletin*, Vol. 33 No.1, pp. 25–36.
- [22] Levenstein, A. (1966). Binary codes capable of correcting deletions, insertions, and reversals, *Soviet Physics Doklady*, Vol. 10 No. 8, pp. 707–710.
- [23] Nguyen, V. T., Sallaberry, C. and Gaio, M. (2013). Mesure de la similarité entre termes et labels de concepts ontologiques, *CORIA 2013*, Neufchâtel, Suisse. pp.415–430. hal-00847528f
- [24] Lindberg, D., Humphrey, B., and McCray, A., (1993) The unified medical language system, in *Methods of Information in Medicine*, Vol. 32 No. 4, pp. 281–291.
- [25] Fellbaum, C. (Ed.). (1998). *WordNet, An Electronic Lexical Database*, MIT Press, Cambridge, MA.
- [26] Ferdinand, M., Zirpins, C. and Trastour, D. (2004). Lifting XML Schema to OWL, in Koch, N., Fraternali, P. and Wirsing, M. (Eds.), *Web Engineering*, Springer-Verlag, Berlin, pp. 354–358.  
[https://doi.org/10.1007/978-3-540-27834-4\\_44](https://doi.org/10.1007/978-3-540-27834-4_44)
- [27] Yahia, N., Mokhtar, S. A. and Ahmed, A. (2012). Automatic generation of OWL ontology from XML data source, *International Journal of Computer Science*, Vol. 9 No. 2, pp. 77–83.
- [28] Hemam, M., and Boufaïda, Z., (2011). Mvp-OWL: A Multi-viewpoints ontology language for the semantic web, *International Journal of Reasoning-based Intelligent Systems*, Inderscience Publishers Ltd. Vol. 3 No. 3, pp. 147–155.  
<https://doi.org/10.1504/ijris.2011.043539>
- [29] Carne, J., Gilleron, R., Lemay, A. and Niehren, J., (2007). Interactive learning of node selecting tree transducers, machine learning, *Machine Learning Journal*, Vol. 66 No.1, pp. 33–67.  
<https://doi.org/10.1007/s10994-006-9613-8>
- [30] Ciravegna, F. and Wilks, Y., (2003). Designing adaptive information extraction for the semantic web in Amilcare, in Handschuh, S. and Staab, S. (Eds.), *Annotation for the Semantic Web*, IOS Press, Amsterdam, pp. 112–127.
- [31] Davulcu, H., Vadrevu, S. and Nagarajan, S., (2005). Ontominer: Automated metadata and instance mining from news websites, *International Journal of Web and Grid Services*, Vol. 1 No. 2, pp. 196–221.  
<https://doi.org/10.1504/ijwgs.2005.008320>
- [32] Euzenat, J. (2002). Eight questions about semantic web annotations, *IEEE Intelligent Systems*, Vol. 17 No. 2, pp. 55–62.  
<https://doi.org/10.1109/5254.999221>
- [33] Petridis, K., Anastasopoulos, D., Saathoff, C., Timmermann, N., Kompatsiaris, Y. and Staab, (2006). M-ontomat-annotizer: Image annotation linking ontologies and multimedia low-level features, in Gabrys, B., Howlett, R.J. and Jain, L.C. (Eds.), *Knowledge-Based Intelligent Information and Engineering Systems*, Springer-Verlag, Berlin, pp. 633–640.  
[https://doi.org/10.1007/11893011\\_80](https://doi.org/10.1007/11893011_80)
- [34] Alani, H., Sanghee, K., Millard, D.E., Weal, M.J., Hall, W., Lewis P.H. and Shadbolt, N.R. (2003). Automatic ontologbased knowledge extraction from web documents, *IEEE Intelligent Systems*, Vol. 18 No. 1, pp. 14–21.  
<https://doi.org/10.1109/mis.2003.1179189>
- [35] Hignette G., Buche P., Dibie-Barthélemy J., Haemmerlé O. (2007) Semantic Annotation of Data Tables Using a Domain Ontology. In: Corruble V., Takeda M., Suzuki E. (eds) *Discovery Science. DS 2007. Lecture Notes in Computer Science*, vol 4755. Springer, Berlin, Heidelberg.  
[https://doi.org/10.1007/978-3-540-75488-6\\_25](https://doi.org/10.1007/978-3-540-75488-6_25)
- [36] Thiam, M. (2010). Annotation sémantique de documents Semi-structuré pour la recherche d'information . PhD thesis, University of Paris-Sud, Gaston Berger, France.
- [37] Borgida, A. and Serafini, L. (2003). Distributed description logics: Assimilating information from

- peer sources, *Journal of Data Semantics*, Vol. 1 No. 1, pp. 153–184.  
[https://doi.org/10.1007/978-3-540-39733-5\\_7](https://doi.org/10.1007/978-3-540-39733-5_7)
- [38] Bouquet P., Giunchiglia F., van Harmelen F., Serafini L., Stuckenschmidt H. (2003) C-OWL: Contextualizing Ontologies. In: Fensel D., Sycara K., Mylopoulos J. (eds) *The Semantic Web - ISWC 2003*. ISWC 2003. Lecture Notes in Computer Science, vol 2870. Springer, Berlin, Heidelberg.  
[https://doi.org/10.1007/978-3-540-39718-2\\_11](https://doi.org/10.1007/978-3-540-39718-2_11)
- [39] Bach., T.L. (2006). *Construction d'un Web sémantique multi-points de vue*. PhD thesis, Mines ParisTech (École des Mines de Paris), Sophia Antipolis, France.
- [40] Bhogal., J., Macfarlane, A., and Smith, P., (2007). A review of ontology based query expansion, *Information Processing & Management*, Vol. 43 No. 4, pp. 866–886.  
<https://doi.org/10.1016/j.ipm.2006.09.003>
- [41] Djezzar., M., and Boufaïda., Z., (2015) Ontological classification of individuals: a multi-viewpoints approach, *International Journal of Reasoning-based Intelligent Systems*, 2015 Vol.7 No.3/4, pp.276–285.  
<https://doi: 10.1504/IJRIS.2015.072954>
- [42] Hemam, M., Djama, O. and Boufaïda, Z. (2010). Une approche d'annotation sémantique à partir d'une ontologie multipoints de vue. *Proceedings of Journées d'Etudes Doctorales*, pp.13–18.
- [43] Horrocks, I., Patel-Schneider, P.F., Boley, H., Tabet, S., Grosof, B., Dean, M.: SWRL: A semantic web rule language combining OWL and RuleML. W3C membersubmission, World Wide Web Consortium.
- [44] Bohring, H. and Auer, S., (2015). Mapping XML to OWL ontologies. In: Jantke, K. P., Fähnrich, K.-P. & Wittig, W. S. (Hrsg.), *Marktplatz Internet: Von e-Learning bis e-Payment*, 13. Leipziger Informatik-Tage (LIT 2005). Bonn: Gesellschaft für Informatik e. V. (S. 147-156).
- [45] Djezzar M., Hemam M., Boufaïda Z. (2012) Ontological Re-classification of Individuals: A Multi-viewpoints Approach. In: Abelló A., Bellatreche L., Benatallah B. (eds) *Model and Data Engineering. MEDI 2012*. Lecture Notes in Computer Science, vol 7602. Springer, Berlin, Heidelberg.  
[https://doi.org/10.1007/978-3-642-33609-6\\_10](https://doi.org/10.1007/978-3-642-33609-6_10)
- [46] Brank, J., Leban, G., and Grobelnik, M., (2018). Semantic Annotation of Documents Based on Wikipedia Concepts. *Informatica (Slovenia) Vol.42 No. 1*, pp. 23–31.