

# Populating an Enterprise Ontology from XML Documents

MamounyAnass<sup>1</sup>, HanouneMostafa<sup>2</sup>

Student, Doctoral Studies: Mathematics, Computer Science and Information Technologies, Hassan II Casablanca University, Faculty of Sciences Ben M'sik, BP 7955 Casablanca, Morocco<sup>1</sup>

Professor, Doctoral Studies: Mathematics, Computer Science and Information Technologies, Hassan II Casablanca University, Faculty of Sciences Ben M'sik, BP 7955 Casablanca, Morocco<sup>2</sup>

**ABSTRACT:** In this work, we propose an approach to populate an ontology named "EntrOnto" describing the internal and external firm environment from a set of XML documents. "EntrOnto" is an ontology without instances whose conception and construction steps were already presented in a previous study. The approach proposed in this paper is organized in two steps: 1) mapping between annotated elements in XML documents and the ontology classes, 2) populating the ontology from the transformed XML data to OWL individuals with the rules previously defined in the first step. The method is successfully applied to a set of international nomenclatures and classifications.

**KEYWORDS:** Populating Ontology, Enterprise, Semantic Web, Competitive Intelligence, Mapping XML to OWL.

## I. INTRODUCTION

The Semantic Web (SW) describes the direct or indirect artificial intervention to generate useful knowledge from web data [1]. We propose to exploit SW techniques to improve the cycle of information collection in Competitive Intelligence (CI) process. The SW model is based on the ontology notion defined as a semantic network which includes a set of concepts describing a given field [2]. In our context, we investigate on the internal and external environment of firms, including factors with different degrees of importance while making a strategic decision. We have described in a previous work the core structure (concepts, relationships, and attributes) of a new ontology named "EntrOnto" without specifying its instances [3]. In this paper, we propose an approach to populate our ontology from a set of XML [4] documents. The first section describes the state of art in this promising field. EntrOnto ontology will be presented in the second section. In the third section, we focus on the novel approach and its steps. Finally, we apply this approach on practical cases to demonstrate its reliability.

## II. STATE OF THE ART

Populating an ontology consists of adding new instances without changing its conceptual structure. Most of the work done on ontology populating is based on analyzing and processing a set of background documents (corpus). Gueffaz M. et al. [5], for instance, have proposed a semi-automatic method to create and enrich ontologies based on the content of the French school programs structured in XML. They have used European Schoolnet Bloom's taxonomy to detect pedagogical skills, and DBpediaSpotLight API to identify named entities. XML documents are then converted into RDF [6] to be mapped to ontologies of different subjects. In the approach proposed by Alec C. et al [7], there is a complementarity between a semi-automatic annotation of a representative cross from the corpus using a reference ontology validated by an expert, and a fully automatic annotation of the rest of the corpus (a product catalog) based on automatic learning mechanisms exploiting the results of the first step and using the linear classifier LIBLINEAR and SVM models. AMARDESH F. et al. [8] have proposed a platform, managed by "Intelligent Topic Manager" (ITM) tool, to populate ontologies from linguistic extractions in the form of conceptual trees made by "Insight Discovery Extractor" (IDE) tool. The mapping between the labelled trees and the ontology concepts is implemented by defining the rules of knowledge acquiring. A linguistic processing chain of tourism web pages converted to XML was proposed by S. Weiser et al. [9] to annotate temporal expressions related to tourism-related objects in order to

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

populate the ontology using knowledge acquisition rules and CA-Manager. According to this work, we deduced that the user intervention in populating ontology process is necessary to have consistent results. Thus, we have identified three important steps for populating: 1) document conversion into a structured language such as XML or RDF, 2) document annotation and instance extraction, 3) mapping between extracted elements and ontology classes. These steps will be considered in our approach in the fourth section.

### III. ENTRONTO

We have demonstrated in a previous work [3] the construction steps of a reference ontology "EntrOnto" that assembles internal and external factors of a firm allowing it to define its strategic needs and decision-making problems. EntrOnto combines: 1) concepts and relations of an ontology developed in the TOVE [10] project describing the internal environment of an organization (resources, structure, roles, communication, goals, etc.), 2) concepts and relations of the immediate external environment of a firm (Micro-Environment) matching to the five forces of Michael Porter [11] model (customers, suppliers, partners, new entrants and substitutes), 3) concepts and relations of the firm external environment (Macro-Environment) which describes the categories of the PESTLE [12] analysis model (Political, Economic, Sociological, Technological, Legal and Ecological). Fig.1 shows the concepts and relationships of EntrOnto, and Fig.2 shows the EntrOntostructure in Protégé tool.

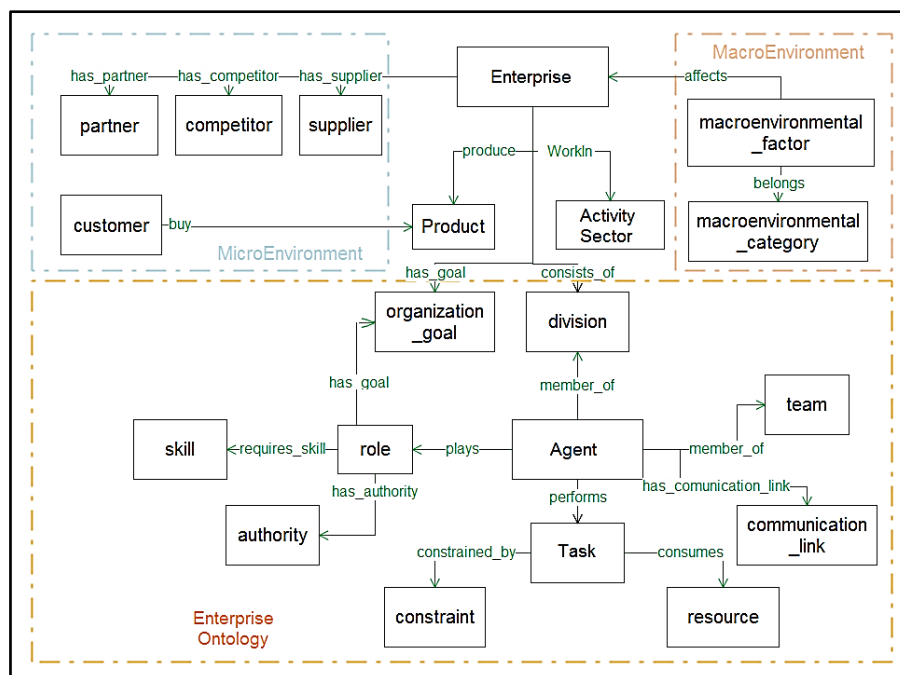


Fig. 1. Diagram of our ontology "EntrOnto".

EntrOnto is a core ontology without instances. Therefore, it is necessary to populate it with individuals in order to be useful for document annotation and for improving the research information pertinence in Competitive Intelligence approaches.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

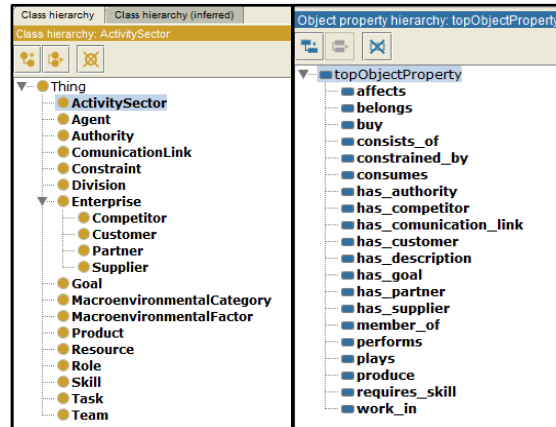


Fig. 2. Classes and Relationships of EntrOnto in Protégé tool.

After an analysis of our ontology, we concluded that there are classes that cannot be populated before knowing the studied firm because they are inconstant. For instance, both classes *Competitor* and *Partner* change from one company to another, and even for a given company a competitor can become a partner or vice versa. However, individuals of *ActivitySector* class, for example, can be performed regardless of the activity sector because it follows standardized norms. We decided to leave to the ontology user the instantiation of the following classes: *Supplier*, *Partner*, *Customer*, *Competitor*, *Division*, *Team*, *Constraint*, *Agent*, and *Task*. We focus in this work on populating the following classes: *ActivitySector*, *Product*, *Role*, *Skill*, *MacroenvironmentalCategory*, *MacroenvironmentalFactor*, *Resource* and *CommunicationLink*.

## IV. PROPOSED APPROACH

The general approach followed in populating our ontology is presented in the Fig. 3. It is composed of two steps: 1) MAPPING and 2) POPULATING. Our approach inputs are: the corpus (set of documents) structured in XML and the basic ontology (EntrOnto) structured in OWL [13]. The output is EntrOnto populated with a set of individuals and structured in OWL.

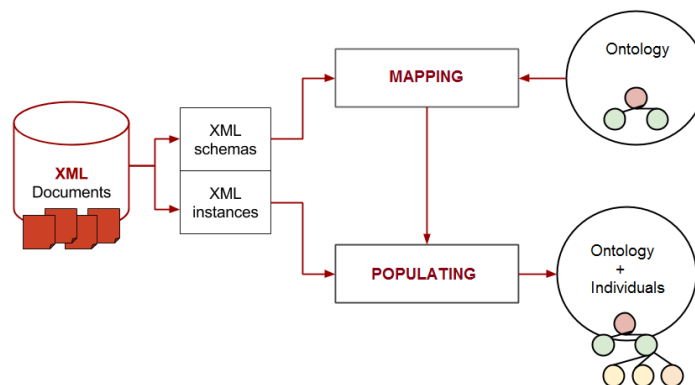


Fig. 3. The proposed approach for populating EntrOnto.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

## 1. MAPPING

Mapping involves four steps : 1) analyzing the XML document structure, 2) defining the XML schema of documents, 3) identifying the components of XML schemas (elements, attributes, simple types, and complex types) that can be labeled, 4) creating mapping rules between the selected components and the corresponding elements of the basic ontology (classes, relationships, attributes). This process is semi-automatic because the user intervention is only required for steps 3) and 4).

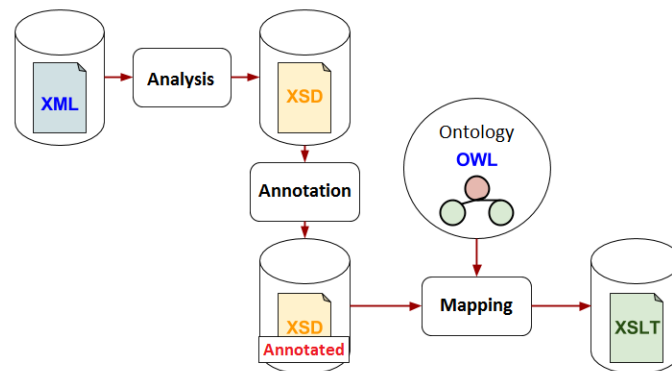


Fig. 4. Mapping process between XML document and EntrOnto.

Fig. 4 shows the procedures involved in the first step of our approach. The XML document analysis should generate a XSD (XML Schema Definition) document which defines the structure and the content type of the XML document. XSD document generation may be performed by a user or automatically using one of the available software tools. In the second step, we mark the useful schematic components that can be used for populating the ontology. We define the mapping rules between the marked XML components and the corresponding EntrOnto elements in an XSLT [14] (eXtensibleStylesheet Language Transformations) file. These rules will be employed in the following step.

## 2. POPULATING

This step can be completely automatic. It invokes the transformation of XML data into OWL individuals through the XSLT file generated in the first step, to define an efficient algorithm which automatically integrates individuals in the ontology and to validate the whole ontology consistency.

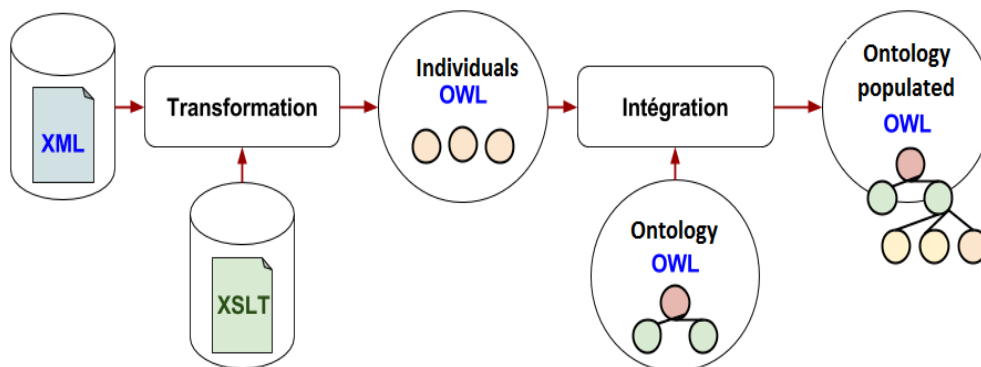


Fig. 5. Populating process of EntrOnto

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

## V. APPROACH EVALUATION

In this section, we apply the proposed approach on a set of international classifications and nomenclatures structured in XML format and downloaded from RAMON(Reference And Management Of BOMs) metadata Eurostat server [15]. Table I shows titles, references and revisions of documents used in approach evaluation.

Table I  
Classifications documents used in approach evaluation

Title	Reference	Revision
International Standard Industrial Classification of All Economic Activities	ISIC_4	Revision 4 (2008)
Central Product Classification	CPC_2	Revision 2 (2008)
International Standard Classification of Occupations	CL_ISCO08	Revision 4 (2008)
European Skills, Competences, Qualifications and Occupations	ESCO-V0	Revision 0 (2014)

We analyze the documents and we mark items to map them to the corresponding objects in our ontology [16]. The Table II shows examples of mapping XSD elements (*Item*, *LabelText*, *PropertyText*) and OWL objects (*Class:ActivitySector*, *DatatypeProperty:description*, *DatatypeProperty:description*).

Table II  
Examples of mapping XSD elements and OWL objects

XSD	OWL
<code>&lt;xs:element name="Item" &gt;</code>	<code>&lt;owl:Classrdf:ID="ActivitySector"/&gt;</code>
<code>&lt;xs:element name="LabelText" type="xs:string"&gt;</code>	<code>&lt;owl:DatatypePropertyrdf:ID="name"&gt;</code>
<code>&lt;xs:element name="PropertyText" type="xs:string"&gt;</code>	<code>&lt;owl:DatatypePropertyrdf:ID="description"&gt;</code>

We use "JXML2OWL Mapper" tool that allows the graphical creation and edition of the correspondence between XML Schema and OWL ontology using JXML2OWL API [17,18], which exports the XSLT transformation file, and transforms the XML instances into OWL instances using the generated XSLT transformation (Fig. 6).

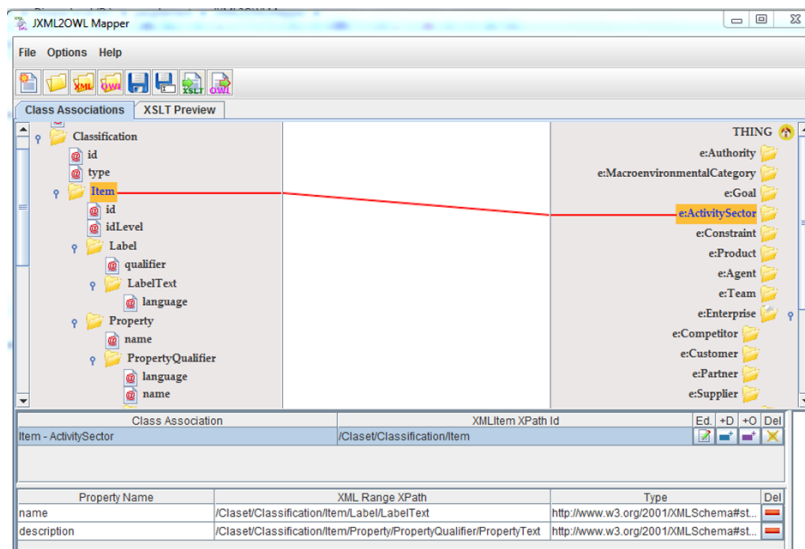


Fig. 6.Mapping ISIC document andEntrOnto classsthrough JXML2OWL tool.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

The results of evaluation approach are shown in Table III. We found 587 individuals for *ActivitySector* class from ISIC\_4 document, 3533 individuals for *Product* class from CPC\_2 document, 606 individuals for *Role* class from CL\_ISCO08 document, and 4876 individuals for *Skill* class from ESCO-V0 document.

Table III  
Resultsofapproach evaluation

XMLDocument	OWL Class	Number of exported individuals
ISIC_4	ActivitySector	587
CPC_2	Product	3533
CL_ISCO08	Role	606
ESCO-V0	Skill	4876

The Individuals found in the documents have been textually integrated in the OWL file of EntrOnto, and then imported to Protégé which is a tool used for validating the ontology consistence trough the inference engine FACT ++ (Fig. 7).

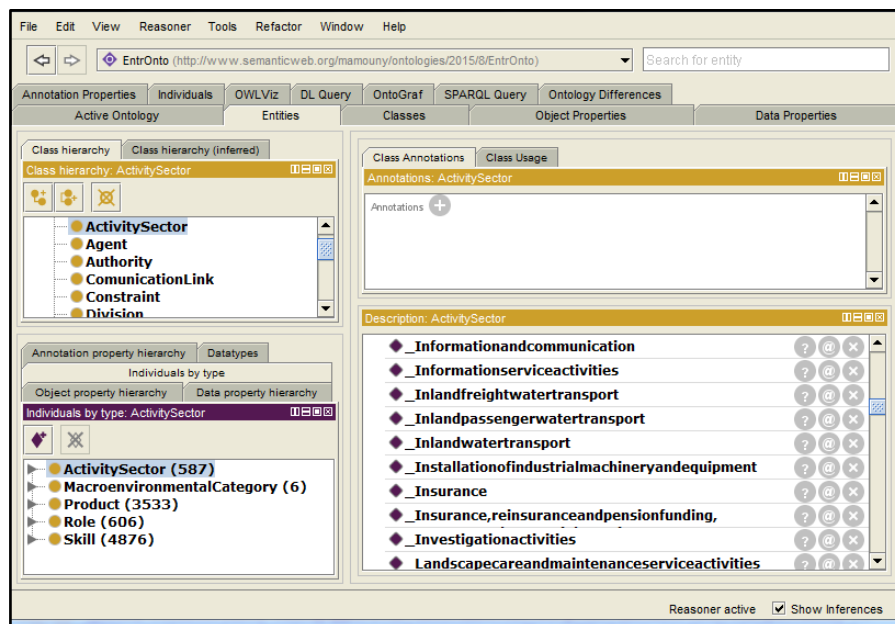


Fig. 7. The imported individuals in Protégé tool.

## VI. CONCLUSION

We developed an approach to populate a firm ontology named "EntrOnto" from XML documents. We created EntrOnto for research and collect of information during the Competitive Intelligence process in the business world. The proposed method is organized into two steps: (1) Mapping between the elements annotated in XML documents and our ontology objects, (2) Populating the ontology from XML data and transforming it into OWL individuals through the rules defined in the first step. The suggested technique was evaluated on a set of economic international classifications and nomenclatures trough JXML2OWL mapping tool. We aim in our future work to apply this procedure to firm web directories, and to demonstrate the added value of using the resulting populated ontology in rationalizing strategic decisions.

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 12, December 2015

## REFERENCES

- [1] Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web, Scientific American, May 2001.
- [2] Gruber, T. Ontology, to appear in the Encyclopedia of Database Systems, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2008.
- [3] Mamouny, A. & al. Using Ontology to Define Decision Problems of Enterprises, International Journal of Engineering Research and Technology Vol. 12 (5), 2013, ISSN 2278 - 0181.
- [4] Biron, P.-V., and Malhotra, A. XML Schema Part 2: Datatypes W3C Recommendation. Technical report, W3C, <http://www.w3.org/TR/2004/REC-xmlschema-2-20041028/>, 2004.
- [5] Gueffaz, M., Deslis, J., et Moissinac, J.-C., Peuplement automatisé d'ontologies par analyse des programmes scolaires, RISE 2014, Nancy, Mars 2014.
- [6] Dan Brickley, R.V. Guha. RDF Schema 1.1. W3C Recommendation. Technical report, W3C, <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>
- [7] Alec, C., Safar, B., Reynaud-Delaître, C., Sellami, Z., Berdugo, U. Peuplement automatique d'ontologie à partir d'un catalogue de produits. Catherine Faron-Zucker. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.87-98, 2014.
- [8] Amardeilh, F., Laublet, P., Minel, J.-L. Annotation documentaire et peuplement d'ontologie à partir d'extractions linguistiques. IC'2005, 2005, pp.100-112, 2005.
- [9] Weiser, S., Coste, M., Amardeilh, F. Chaîne de traitement linguistique : du repérage d'expression temporelle au peuplement d'une ontologie de tourisme. TALN 2009 – Session posters, Senlis, 24-26 juin, 2009.
- [10] TOVE Ontology Project. Available: <http://www.eil.utoronto.ca/enterprisemodelling/tove/>
- [11] Porter, M.E. Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press, 2008.
- [12] TEAM FME, PESTLE ANALYSIS, Strategy Skills, ISBN 978-1-62620-998-5. [www.free-management-ebooks.com](http://www.free-management-ebooks.com), 2013.
- [13] Bechhofer, S. Web Ontology Language (OWL) Reference version 1.0. W3C. Technical report, W3C, <http://w3.org/TR/owl-ref/>, 2004.
- [14] Clark, J. XSL Transformations (XSLT). Technical report, W3C, <http://www.w3.org/TR/xslt>, 1999.
- [15] Site officiel de RAMON : [http://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP\\_PUB\\_WELC](http://ec.europa.eu/eurostat/ramon/index.cfm?TargetUrl=DSP_PUB_WELC)
- [16] Anicic, N., Ivezic, N. and Marjanovic, Z.: Mapping XML Schema to OWL, Enterprise Interoperability, Springer London, 2007.
- [17] Rodrigues, T., Rosa, P. and Cardoso, J.: Mapping XML to Existing OWL ontologies, International Conference WWW/Internet, 2006.
- [18] <http://jxml2owl.projects.semwebcentral.org/>
- [19] Ferdinand, M., Zirpins, C., & Trastour, D., Lifting XML Schema to OWL. Web Engineering Lecture Notes in Computer Science Volume 3140, pp. 354-358, 2004.
- [20] Ghawi, R., & Cullot, N., Building Ontologies from XML Data Sources. DEXA '09. 20th International Workshop on Database and Expert Systems Application, pp. 480-484, 2009.
- [21] Cruz, C., Nicolle, C. Graphe-Based Rules For XML Data Conversion to OWL Ontology. WEBIST 2010, Apr 2010, Valencia, Spain. INSTICC Press, pp.175-178, isbn: 978-989-674-025-2, 2010.
- [22] Yahia, N., Mokhtar, S. A., Ahmed, A. Automatic generation of OWL ontology from XML data source. In International Journal of Computer Science Issues (IJCSI) 9(2), 2012.
- [23] He Tan, Georges Barakat, Vladimir Tarasov: Translating XML Models into OWL Ontologies for Interoperability of Simulation Systems. BIR Workshops 2015: 116-123, 2015.