

# Peuplement d'une ontologie des entreprises à partir des documents XML

MAMOUNY Anass<sup>1</sup>, HANOUNE Mostafa<sup>2</sup>

<sup>1</sup> Faculté des sciences Ben M'Sik, Université Hassan II de Casablanca, Maroc,  
anass.mamouny@gmail.com

<sup>2</sup> Faculté des sciences Ben M'Sik, Université Hassan II de Casablanca, Maroc,  
mhanoune@gmail.com

**Résumé-** Dans le présent article, nous proposons une approche pour peupler une ontologie nommée « EntrOnto » décrivant l'environnement intérieur et extérieur des entreprises à partir d'un ensemble de documents structurés en XML. « EntrOnto » est une ontologie sans individus, les étapes de sa conception et de sa construction ont été déjà présentées dans un article antérieur. L'approche proposée dans cet article est organisée en deux étapes : (1) d'Annotation des documents XML et mapping entre les éléments annotés et les classes de l'ontologie, (2) Peuplement de l'ontologie à partir des données XML transformées en des individus OWL grâce aux règles définies dans la première étape. Le travail a été appliqué sur un ensemble de nomenclatures et de classifications internationales.

**Mots-clés :** Peuplement d'Ontologie, Entreprise, Web Sémantique, Intelligence Économique, Mapping XML et OWL

## INTRODUCTION

Selon Tim Berners-Lee, le Web Sémantique (WS) est «un web de données qui peuvent être traitées directement et indirectement par des machines pour aider leurs utilisateurs à créer de nouvelles connaissances». Nous proposons d'exploiter les techniques du WS pour améliorer le cycle de la collecte d'informations dans la démarche de l'Intelligence Économique (IE) adoptée par les entreprises visant à être plus compétitives. Le modèle du WS se base essentiellement sur la notion d'Ontologie définie par un réseau sémantique qui regroupe un ensemble de concepts décrivant un certain domaine [1]. Le domaine décrit dans le contexte de notre travail est l'environnement intérieur et extérieur d'une entreprise, y compris les facteurs considérés globalement ou partiellement lors de la prise d'une

décision stratégique. Nous avons déterminé dans un article précédent la structure de base (les concepts, les relations, et les attributs) d'une nouvelle ontologie sans instances nommée «EntrOnto» [2]. Dans le présent article, nous proposons une approche pour peupler et instancier notre ontologie à partir d'un ensemble de documents structurés en XML [3]. La première section décrit l'état d'art des travaux antérieurs. L'ontologie EntrOnto sera présentée dans la deuxième section. Dans la troisième section, nous focalisons sur l'approche proposée et ses étapes. Enfin, nous appliquons cette approche sur des cas pratique pour démontrer sa fiabilité.

## ETAT DE L'ART

Peupler une ontologie consiste à y ajouter de nouvelles instances sans en changer la structure conceptuelle. Divers travaux de peuplement d'ontologie ont été proposés sur plusieurs domaines d'application et la plupart de ces approches se basent sur l'analyse et le traitement d'un ensemble de documents de base (corpus). Par exemple M. Gueffaz et al. [4] ont proposé une méthode semi-automatique de création et d'enrichissement d'ontologies en s'appuyant sur le contenu des programmes scolaires français structurés en XML. Ils se sont servis de la taxonomie de Bloom proposée par l'European Schoolnet pour la détection des capacités pédagogiques, et de l'API DBpedia SpotLight pour la détection des entités nommées. Les documents XML ont été ensuite transformés en RDF [5] afin d'être mappés vers les ontologies de base

des différentes matières. Dans l'approche proposée par C. Alec et al [6], on trouve une complémentarité entre une étape semi-automatique d'annotation, validée par un expert du domaine, portant sur un échantillon représentatif du corpus en utilisant une ontologie de domaine de référence, et une étape entièrement automatique d'annotation du reste du corpus (un catalogue de produits) basée sur des mécanismes d'apprentissage automatiques exploitant les résultats de la première étape en utilisant le classifieur linéaire *LIBLINEAR* et les modèles SVM. F. Amardeilh et al. [7] ont proposé une plateforme qui permet de peupler une ontologie gérée par l'outil « *Intelligent Topic Manager* » (ITM) à partir des extractions linguistiques sous forme d'arbres conceptuels effectués par l'outil « *Insight Discoverer Extractor* » (IDE). Le Mapping entre les étiquettes des arbres et les concepts de l'ontologie s'effectue grâce à la définition des règles d'acquisition de connaissance. Une chaîne de traitement linguistique des pages web touristiques converties en XML a été proposée par S. Weiser et al. [8] pour annoter les expressions temporelles liées aux objets touristiques, afin de peupler l'ontologie de domaine à l'aide des règles d'acquisition de connaissance et l'outil CA-Manager. D'après les travaux cités, nous avons déduit que l'intervention de l'utilisateur dans le peuplement d'ontologie est nécessaire afin d'avoir des résultats cohérents. Ainsi, nous avons retenu trois étapes importantes pour le peuplement: (1) la conversion des documents en un langage structuré tel que XML ou RDF, (2) l'annotation des documents et l'extraction des instances, (3) le mapping entre les éléments extraits et les classes de l'ontologie. Nous prenons en considération ces étapes pour notre approche proposée dans la quatrième section.

### L'ONTOLOGIE ENTRONTO

Nous avons montré dans un article précédent [2] les étapes de la construction d'une ontologie de base sous le nom « EntrOnto » qui rassemble les facteurs internes et externes d'une entreprise permettant de définir ses besoins stratégiques et ses problèmes décisionnels. EntrOnto regroupe les concepts et les relations d'une ontologie développée dans le cadre de projet TOVE [9] décrivant l'environnement interne d'une organisation (les ressources, la structure, les rôles, les moyens de communication, les objectifs, etc.), les concepts et les relations de l'environnement externe immédiat d'une entreprise (Microenvironnement) décrivant les cinq forces de modèle de Michael Porter [10] (les clients, les fournisseurs, les partenaires, les nouveaux entrants, et les produits de substitution), et les concepts et les relations de l'environnement extérieur global d'une entreprise (Macro-environnement) décrivant les catégories du modèle d'analyse PESTEL [11] (Politique, Economique, Sociologique, Technologique, Ecologique, et Légal). Le diagramme de la figure (Fig.1) montre les concepts et les relations d'EntrOnto. La figure (Fig.2) présente la hiérarchie des concepts et des relations d'EntrOnto dans l'outil Protégé.

## Revue de l'Entrepreneuriat et de l'Innovation

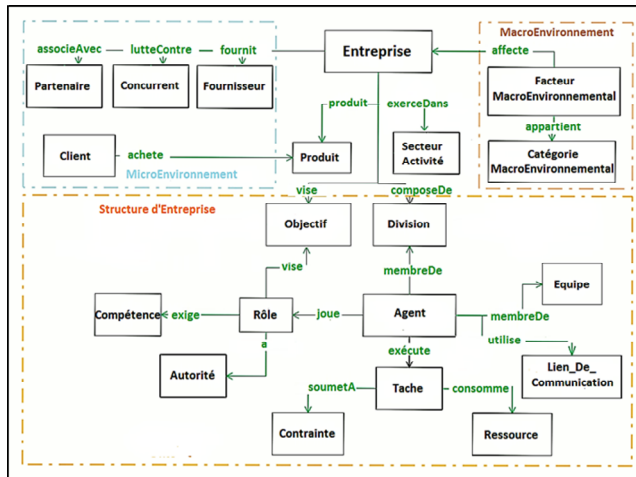


Fig. 1 Diagramme de notre ontologie « EntrOnto ».

EntrOnto est une ontologie de base sans instances. Il est donc nécessaire de la peupler avec des individus pour pouvoir l'utiliser dans l'annotation des documents et l'amélioration de la pertinence de la recherche d'information au cours de la démarche d'Intelligence Economique.

les individus de ces deux classes avant de connaître l'entreprise utilisatrice de l'ontologie. En revanche, on peut proposer des individus pour la classe *SecteurActivite* sans connaître l'entreprise utilisatrice de l'ontologie puisque les secteurs d'activités économiques sont standards et connus. On a décidé alors de laisser aux Veilleur de l'entreprise utilisatrice de l'ontologie la tâche d'instanciation des classes suivantes: *Fournisseur*, *Partenaire*, *Client*, *Concurrent*, *Division*, *Equipe*, *Contrainte*, *Agent*, et *Tache*. Nous nous occupons dans cet article du peuplement des classes suivantes : *SecteurActivite*, *Produit*, *Rôle*, *Compétence*, *CategorieMacroenvironnemental*, *FacteurMacroenvironnemental*, *LienDeCommunication*, et *Ressource*.

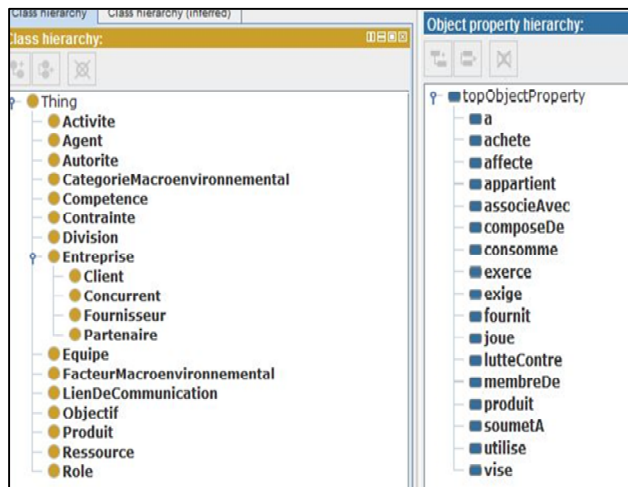


Fig.2 Classes et Relations d'EntrOnto sur l'outil Protégé.

Après une analyse de notre ontologie, on a conclu qu'il y a des classes qu'on ne peut pas instancier car ils sont liés strictement au contexte de l'entreprise. Par exemple, les deux classes *Concurrent* et *Partenaire* changent d'une entreprise à une autre, et même pour une entreprise donnée un concurrent peut devenir un partenaire ou l'inverse. Donc on ne peut pas déterminer

### APPROCHE PROPOSÉE

L'approche générale suivie au cours du peuplement de notre ontologie est présentée dans la figure (Fig.3). Elle est composée principalement de deux étapes : (1) MAPPING, et (2) PEUPLEMENT. Les inputs de notre approche sont : le corpus (l'ensemble des documents) structuré en XML et l'ontologie de base (EntrOnto) structurée en OWL [12]. L'output est l'ontologie de base (EntrOnto) peuplée avec un ensemble d'individus et structurée avec le même langage OWL.

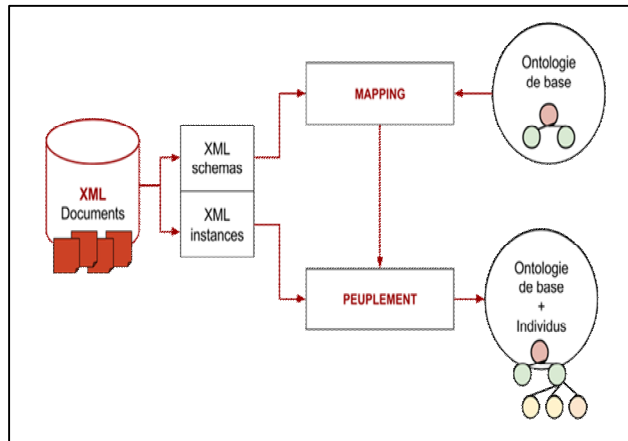


Fig.3 L'Approche proposée pour peupler l'ontologie EntrOnto.

### 1. MAPPING

La première étape consiste à (a) analyser la structure des documents XML, (b) définir les schémas XML des documents, (c) identifier les composants des schémas XML (éléments, attributs, types simples et types complexes) à marquer, et (d) créer des règles de mapping entre les composants marqués et les composants correspondants de l'ontologie de base (classes, relations, attributs). Cette procédure est semi-automatique puisque l'intervention de l'utilisateur est obligatoire pour les étapes (c) et (d).

La figure (Fig.4) montre les sous-étapes de la première étape de notre approche. L'analyse de document XML doit générer un document XSD (XML Schema Definition) permettant de définir la structure et le type du contenu du document XML. La génération du document XSD peut être réalisée par un utilisateur ou automatiquement par un

logiciel.

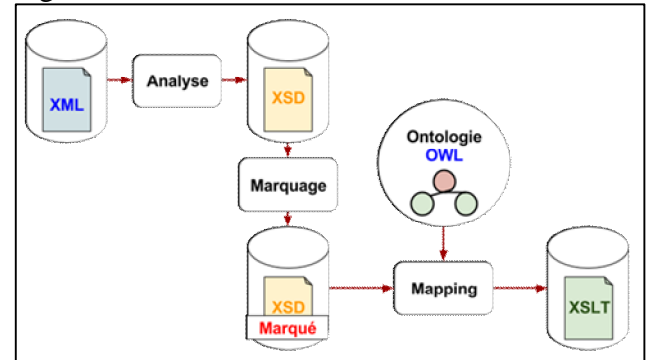


Fig.4 La procédure de Mapping entre un document XML et l'ontologie EntrOnto.

Dans un deuxième temps, nous marquons les composants schématiques utiles qui peuvent servir pour le peuplement de notre ontologie. Nous définissons après les règles de mapping entre les composants XML marqués et les composants OWL correspondants de notre ontologie de base EntrOnto dans un fichier XSLT (eXtensible Stylesheet Language Transformations) [13] qui sera utilisé dans la prochaine étape.

### 2. PEUPLEMENT

Cette étape peut être totalement automatique. Elle consiste à transformer les données XML en individus OWL grâce au fichier XSLT généré dans la première étape, à définir un algorithme pour intégrer les individus automatiquement au sein de l'ontologie, et à valider la cohérence de l'ensemble de l'ontologie.

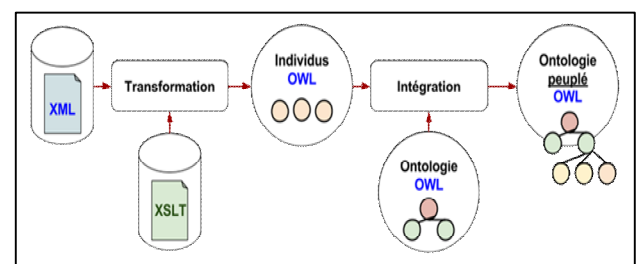


Fig.5 La procédure de peuplement de l'ontologie EntrOnto.

### EVALUATION DE L'APPROCHE

Dans cette section, on va appliquer l'approche proposée sur un ensemble de nomenclatures et de classifications internationales révisées et traduites à la langue française par l'Institut National de la Statistique et des Etudes Economiques [14] (INSEE) sous l'égide de la Commission Nationale des Nomenclatures Economiques et Sociales (CNNES) du Conseil National de l'Information Statistique (CNIS).

Tableau I

Références des Nomenclatures et Classifications utilisées dans l'évaluation de notre approche

Intitulé	Reference Européenne	Reference International
Nomenclature statistique des activités économiques dans la Communauté européenne	NACE_REV2	CITL_REV2
Classification statistique des produits associée aux activités dans la Communauté économique européenne	CPA_2008	CPC_VER2.1
Classification internationale des types des professions	ISCO_88_CO	CITP_08
Classification européenne des aptitudes/compétences, certifications et professions	ESCO	ESCO

Les documents cités dans le tableau I sont structurés au format XML et téléchargés depuis RAMON (Reference And Management Of Nomenclatures), le serveur de métadonnées d'EUROSTAT [16]. On a analysé les documents et on a déterminé les éléments qu'il faut marquer pour les mapper avec les objets correspondants de notre ontologie [17]. Le

tableau II montre des exemples de mapping entre des éléments XSD et des objets OWL.

Tableau II

Exemple de mapping entre les éléments XSD et les objets OWL

XSD	OWL
<code>&lt;xs:element name="Item" &gt;</code>	<code>&lt;owl:Class rdf:ID="SecteurActivite"/&gt;</code>
<code>&lt;xs:element name="LabelText" type="xs:string"&gt;</code>	<code>&lt;owl:DatatypeProperty rdf:ID="description"&gt;</code>
<code>&lt;xs:element name="PropertyText" type="xs:string"&gt;</code>	<code>&lt;owl:DatatypeProperty rdf:ID="nom"&gt;</code>

On a utilisé l'outil « JXML2OWL Mapper » qui permet de créer et de modifier graphiquement la correspondance entre un schéma XML et une ontologie OWL en utilisant l'API JXML2OWL [18,19], d'exporter la transformation XSLT dans un fichier, et de transformer les instances XML en des instances OWL en utilisant la transformation XSLT générée précédemment (Fig.6). Les résultats obtenus sont affichés dans le Tableau III. Les individus trouvés dans les documents ont été intégrés textuellement dans le fichier OWL d'EntrOnto, puis importés dans l'outil Protégé pour vérifier leur cohérence grâce au moteur d'inférence FaCT++ (Fig.7).



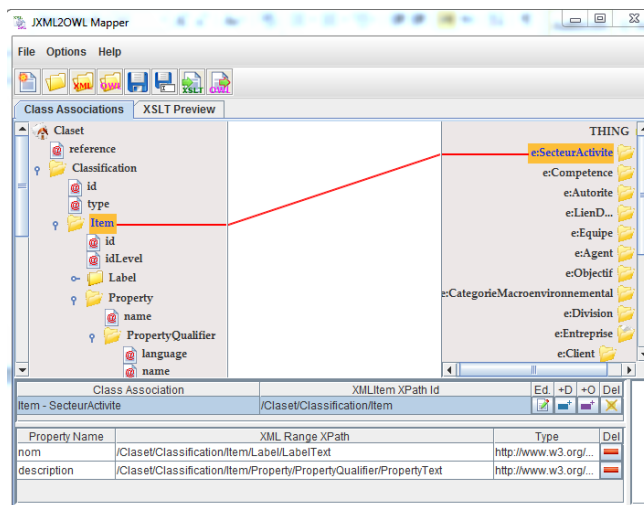


Fig.6 Mapping entre le document NACE et l'ontologie EntrOnto avec l'outil JXML2OWL.

Tableau III

Résultats de l'évaluation de l'approche proposée

Document XML	Classe OWL correspondante	Nombre d'individus exportés
NACE_REV2	SecteurActivite	843
CPA_2008	Produit	4273
ISCO_88_CO	Role	490
ESCO	Competence	4877

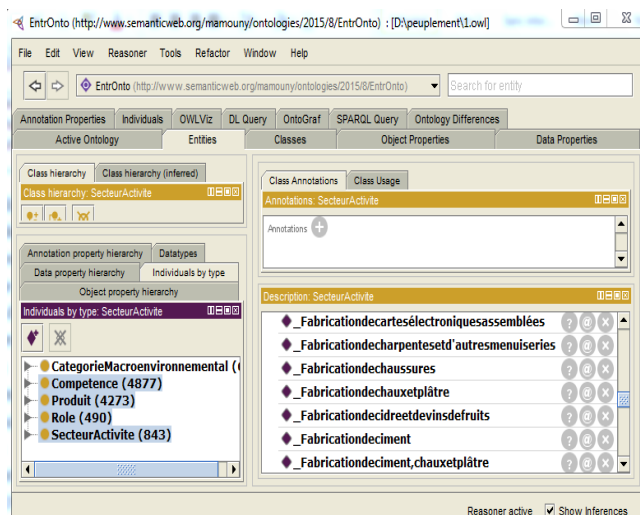


Fig.7 Les individus importés dans PROTEGE.

## CONCLUSION

Dans cet article, nous avons proposé une approche pour peupler une ontologie d'entreprises nommée « EntrOnto » à partir des documents structurés en XML. Nous avons créé EntrOnto afin de l'utiliser pour la recherche et la collecte d'informations au cours de la démarche de l'Intelligence Économique au sein du monde d'entreprise.

L'approche proposée est organisée en deux étapes : (1) Mapping entre les éléments annotés dans les documents XML et les objets de notre l'ontologie, (2) Peuplement de l'ontologie à partir des données XML transformées en des individus OWL grâce aux règles définies dans la première étape. L'approche a été évaluée sur un ensemble de nomenclatures et de classifications économiques, standards, et internationales grâce à l'outil JXML2OWL MAPPER.

Nous visons dans nos futurs travaux d'appliquer cette approche sur les annuaires web des entreprises, et à montrer la valeur ajoutée de l'utilisation de notre ontologie peuplée en matière de rationalisation des décisions stratégiques d'une entreprise.

## Références

- [1] Gruber, T. Ontology, to appear in the Encyclopedia of Database Systems, Ling Liu and M. Tamer Özsu (Eds.), Springer-Verlag, 2008.
  - [2] 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.
  - [3] 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.
- Gueffaz, M., Deslis, J., et Moissinac, J.-C., Peuplement automatisé d'ontologies par analyse des programmes scolaires, RISE 2014, Nancy, Mars 2014.
- Dan Brickley, R.V. Guha. RDF Schema 1.1. W3C Recommendation. Technical report, W3C, <http://www.w3.org/TR/2014/REC-rdf-schema-20140225/>
- 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.
- 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.
- 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.
- TOVE Ontology Project. Available: <http://www.eil.utoronto.ca/enterprisemodelling/tove/>
- Porter, M.E. Competitive Strategy: Techniques for Analyzing Industries and Competitors. Free Press, 2008.
- TEAM FME, PESTLE ANALYSIS, Strategy Skills, ISBN 978-1-62620-998-5, [www.free-management-ebooks.com](http://www.free-management-ebooks.com), 2013.
- Bechhofer, S. Web Ontology Language (OWL) Reference version 1.0. W3C. Technical report, W3C, <http://w3.org/TR/owl-ref/>, 2004.
- Clark, J. XSL Transformations (XSLT). Technical report, W3C, <http://www.w3.org/TR/xslt>, 1999.
- Site officiel de l'INSEE : <http://www.insee.fr/fr/>
- Site officiel de CNIS : <http://www.cnis.fr/cms>
- 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)
- Anicic, N., Ivezic, N. and Marjanovic, Z.: Mapping XML Schema to OWL, Enterprise Interoperability, Springer London, 2007.
- Rodrigues, T., Rosa, P. and Cardoso, J.: Mapping XML to Existing OWL ontologies, International Conference WWW/Internet, 2006.
- <http://jxml2owl.projects.semwebcentral.org/>
- Ferdinand, M., Zirpins, C., & Trastour, D., Lifting XML Schema to OWL. Web Engineering Lecture Notes in Computer Science Volume 3140, pp. 354-358. 2004.
- 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.
- 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.
- 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.
- He Tan, Georges Barakat, Vladimir Tarasov: Translating XML Models into OWL Ontologies for Interoperability of Simulation Systems. BIR Workshops 2015: 116-123, 2015.