



Une cartographie de domaine pour guider la construction d'une ressource sémantique

Sarra Ben Abbès, Haïfa Zargayouna, Adeline Nazarenko

► To cite this version:

Sarra Ben Abbès, Haïfa Zargayouna, Adeline Nazarenko. Une cartographie de domaine pour guider la construction d'une ressource sémantique. 23es Journées francophones d'Ingénierie des Connaissances, Jun 2012, Paris, France. pp.100-103. hal-00703173

HAL Id: hal-00703173

<https://hal.science/hal-00703173>

Submitted on 1 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Une cartographie de domaine pour guider la construction d'une ressource sémantique

Sarra Ben Abbès, Haïfa Zargayouna, Adeline Nazarenko

LIPN - UNIVERSITÉ PARIS 13 & CNRS (UMR 7030)
prenom.nom@lipn.univ-paris13.fr

Résumé : Dans cet article, nous proposons une méthode d'aide à l'ingénieur de la connaissance pour construire un modèle sémantique (terminologie ou ontologie) pour un domaine particulier. Cette méthode repose sur l'utilisation de ressources sémantiques disponibles pour représenter une vue synthétique du domaine à modéliser. La méthode proposée se fonde sur les phases suivantes : modularisation, alignement et raisonnement. Le texte sert de support aux différentes phases.

Mots-clés : ressource sémantique, modularisation, alignement, raisonnement

Construire un modèle sémantique nécessite soit : (i) partir de rien, ou (ii) réutiliser les modèles de connaissances mis à disposition (Abadie & Mustière, 2010; Cimiano & Völker, 2005), ce que nous appelons ici *ressources sémantiques*. Notre travail s'inscrit dans la deuxième catégorie où le but est de réutiliser au mieux les ressources existantes. Le but de ce travail n'est pas de produire de nouvelles connaissances mais de construire une vue synthétique qui peut servir d'amorce au travail de modélisation.

Un des défis de ce travail est lié à l'hétérogénéité de ressources sémantiques disponibles. Elles sont de différentes natures ; elles manipulent des connaissances terminologiques (thésaurus et dictionnaire) ou conceptuelles (ontologie et taxonomie). Ces ressources couvrent plus au moins partiellement un domaine de spécialité. Il existe des ressources qui contiennent des notions générales sur un domaine d'une activité spécifique (ex. Eurovoc) et d'autres qui sont spécifiques à un sujet donné d'un domaine particulier (ex. l'ontologie Kaon décrivant les plantes : fleurs, couleur, longueur, etc). Le degré de granularité de la description du domaine est aussi différent d'une

ressource à l'autre.

Nous construisons une *cartographie de domaine* représentant un ensemble de liens entre les connaissances des différentes ressources recensées. La méthode proposée s'appuie sur un texte qui décrit le domaine à modéliser. Une première phase d'ancrage permet de lier les ressources au texte. Le principe de cette phase est de projeter les éléments des ressources sur le texte afin d'obtenir des éléments ancrés (éléments présents dans le texte). Notre méthode se fonde sur la définition d'un modèle de cartographie qui donne une représentation homogène des connaissances de ressources qui la composent et de la richesse des liens existants entre ces ressources. Notre contribution s'appuie sur les phases suivantes :

1. **Modularisation** : le but de la modularisation est de faciliter l'exploitation des connaissances d'une ressource par l'ingénieur de la connaissance sans le noyer par des connaissances périphériques. Pour ce faire, nous nous appuyons sur les éléments ancrés et construisons des îlots de connaissances (sous-blocs ou portions d'une ressource sémantique) qui délimitent le champs sémantique. Le recours au texte sert à déterminer les connaissances qui se trouvent autour des éléments ancrés tels que les fils, les frères.
2. **Alignement** : la phase d'alignement s'appuie sur des correspondances entre les éléments des ressources sémantiques. Nous nous intéressons aux correspondances de type 1 : 1 (les liens *bleus* dans la figure 1). Le texte sert de support pour : (i) vérifier les alignements réalisés, et (ii) enrichir les alignements entre les ressources sémantiques. L'exploitation du texte permet de raisonner sur sa structure en s'appuyant sur des éléments de contexte (qui peut être la phrase ou le paragraphe).
3. **Raisonnement** : cette phase permet de réaliser des calculs aussi bien sur les correspondances que sur les îlots des ressources obtenus par la modularisation. Les correspondances entre les éléments de ressources permettent d'exhiber des correspondances entre les îlots de connaissances (les liens *verts* dans la figure 1). Les correspondances peuvent aussi permettre de revoir les contours des îlots de connaissances en prenant en compte des informations liées au texte (distribution, couverture).

La cartographie de domaine obtenue en sortie de cette méthode, décrit un ensemble de correspondances entre les entités (éléments et îlots de

connaissances) des ressources avec des caractéristiques spécifiques liées au texte d'acquisition.

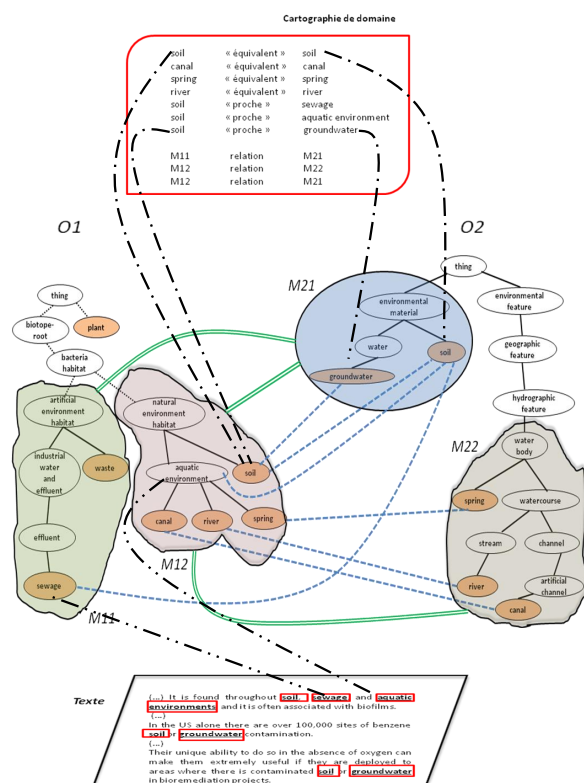


FIGURE 1 – Exemple d'une cartographie de domaine

Références

- ABADIE N. & MUSTIÈRE S. (2010). Constitution et exploitation d'une taxonomie géographique à partir des spécifications de bases de données. *Revue Internationale de Géomatique*, **20**, 145–174.
- CIMIANO P. & VÖLKER J. (2005). Text2onto - a framework for ontology learning and data-driven change discovery.