

Elaboration et utilisation d'une base de connaissances d'un domaine technique.

Nicolas Faure*, René-Michel Faure**

* Centre de recherches Magellan, IAE, Université Lyon 3
mkd.onto@gmail.com

** Ecole Nationale des Travaux Publics de l'Etat, Lyon
ramcesh.rm@orange.fr

Ce poster rend compte d'une entreprise d'élaboration d'un système de représentation des connaissances pour le domaine géotechnique.

1. Caractérisation du domaine

Ce domaine se caractérise en effet par une grande disparité des données techniques (données largement inconnues ou approchées, liées aux caractéristiques physiques des sols) et une forte propension pour la documentation du domaine à être d'ordre descriptif, le lecteur devant raisonner par analogie pour savoir si l'article sera pertinent pour son problème.

2. Objectif poursuivi

Trouver dans la base de données, les articles qui contiennent des phrases qui aideront l'utilisateur à maîtriser son thème de recherche.

3. Source des données

Les données d'origine sont extraites de corpus documentaires rassemblés par les ingénieurs du domaine. Les corpus ainsi constitués représentent plusieurs années des travaux d'une communauté transcrits dans des congrès annuels (AFTES ou AITES en l'occurrence).

4. Originalité de la démarche

Pour répondre à notre problématique, nous avons d'abord envisagé une ontologie de domaine formalisée en OWL, mais nous nous sommes heurtés à plusieurs difficultés :

- La multiplicité des spécialités géotechniques induit une multiplicité d'emplois lexicaux spécialisés et un coût de normalisation assez élevé.
- Le raisonnement par catégories aristotéliennes correspond au final assez peu au raisonnement usuellement employé par les experts du domaine, qui raisonnent plutôt par prototypes et proximité sémantique.
- Il existe aussi dans la profession une résistance diffuse à l'utilisation d'un formalisme de représentation étranger au mode de pensée des experts, et plus encore à toute tentative de normaliser ces modes de pensée.

Donc, le projet est construit sur une base purement lexicale et informelle, sans formalisme de représentation nécessitant spécification ou formalisation des données de départ.

5. Traitement des données

Les données lexicales sont extraites par OCR et traitement morphosyntaxique, permettant de lister les syntagmes signifiants du corpus. Ces syntagmes sont ensuite manuellement classés par les utilisateurs selon des relations simples (hyponymie/hypéronymie, voisinage) au sein de cartes. Ces cartes représentent généralement une thématique ou une spécialité. En général, elles n'excèdent pas 800 syntagmes et respectent quelques contraintes simples. Ces cartes ainsi constituées sont stockées dans une base de données, aident à l'interrogation et sont surtout un mode s'appropriation du système par les utilisateurs. Une analyse des textes permet de détecter les ensembles de phrases déductives de type ; si...alors, qui traduisent en général la connaissance des phénomènes.

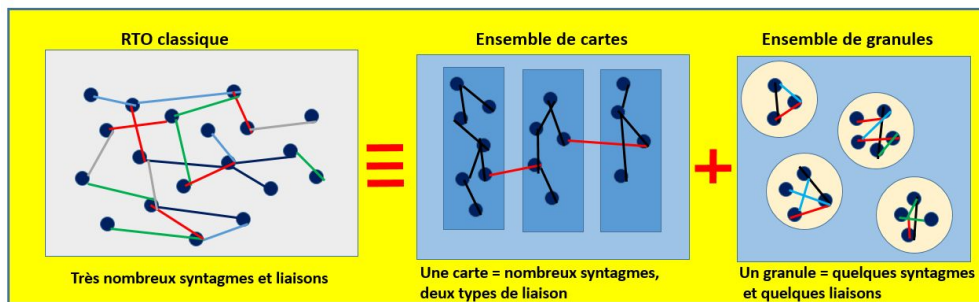
6. Interrogation de la base de connaissances

L'utilisateur peut ensuite interroger la base selon plusieurs modes ; via les noms d'auteur ou via un ensemble de syntagmes constituant une. La recherche des phrases d'articles permet des comparaisons par ressemblance et précise l'origine des informations. De même, les articles peuvent être comparés entre eux et classés par degré de ressemblance.

7. L'avenir

Les essais ont été menés auprès d'experts du domaine en situation réelle fournissent des résultats complets et surtout moins coûteux en temps.

Il est également prévu d'intégrer une connaissance plus complexe dans le système, sous la forme du granule de connaissances, représentation semi-formelle de connaissances



axiomatiques. L'objectif est d'en arriver à une forme de RTO informelle qui permette d'articuler les connaissances entre elles et de fournir des réponses plus pertinentes.

Plusieurs familles de congrès (JNGG, ITA, AFTES) représentant plusieurs milliers d'articles ont été mis en base. Cette application test, ainsi qu'une bibliographie complète, est en ligne à : www.pentes-tunnels.eu/MKD/MKD_en_ligne