

Introduction générale

La construction des bases de données lexicales à partir de textes constitue un sous-domaine à part entière de l'ingénierie des ontologies. Ces bases peuvent servir à titre d'exemples :

- Pour la traduction automatique,
- Pour la recherche de documents.
- Pour les systèmes de questions / réponses.
- Pour la génération de textes.
- Pour la désambiguïsation et la compréhension de textes.

Les travaux dans ce domaine existent déjà pour les langues comme l'anglais ou le français, malheureusement pour l'arabe les choses ne font que commencer. Nous essayons dans ce projet de mémoire de fin d'étude de proposer des outils pour la construction de bases de données lexicales pour la langue arabe.

Cette problématique constitue un enjeu important pour le Traitement Automatique des Langues Naturelles (TALN). En effet les systèmes de TALN ne peuvent être efficaces que s'ils reposent sur des ressources lexicales bien construites pour le domaine et l'application concernés.

Ce mémoire est composé de trois chapitres ; après une introduction générale, on expose dans le premier chapitre le domaine des ontologies et les méthodologies de leurs constructions.

Le deuxième chapitre est consacré entièrement à la description de la structure de l'ontologie lexicale « WordNet » considéré comme modèle de référence pour notre base de données lexicale. L'accent est mis surtout sur la représentation des connaissances.

Dans le troisième chapitre nous avons présenté les grandes lignes de notre implémentation de WordNet pour l'Arabe « AraWord ». Le détail de notre base lexicale est ainsi exposé.

Nous terminons notre mémoire par une conclusion générale, dans laquelle on expose les résultats obtenus ainsi que quelques perspectives pour la continuation de ce travail.