

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Monsieur CRESTAN Eric soutiendra une thèse
Le Jeudi 15 décembre 2005 à 14 h 00

—
**Amphithéâtre
IUP GMI - AVIGNON**

SPÉCIALITÉ : Informatique

Titre de la thèse : *Prise en Compte de Niveaux Sémantiques pour la Détermination du Sens et la Recherche d'Information.*

Membres du jury :

M. MARTEAU Pierre-François, Professeur, Université de Bretagne Sud.
M. GUENTHNER Franz, Professeur, Ludwig-Maximilians-Universität München (Allemagne).
M. EL BEZE Marc, Professeur, Laboratoire d'Informatique d'Avignon, I.U.P., Avignon.
Mme NAZARENKO Adeline, Professeur, Laboratoire LIPN – CNRS UMR, Université Paris-Nord.
M. LOUPY Claude, Docteur, Laboratoire indépendant, Paris.

Résumé de la thèse :

L'objectif de cette thèse est de montrer que la prise en compte de niveaux sémantiques et pragmatiques permet d'accroître les performances des outils traditionnels de recherche d'information. Dans le cadre des évaluations des systèmes de désambiguïssations sémantiques SENSEVAL-1 et 2, les expériences menées montrent que l'emploi d'un niveau sémantique grossier permet d'accroître la précision pour la désambiguïssation des sens fins.

Dans un second temps, une approche non-supervisée a été mise en place, basée sur un dictionnaire thématique existant. L'hypothèse de non-ambiguïté des cooccurrences de termes d'une même phrase, autorise l'inférence de vecteurs sémantiques dont les composantes sont liées aux 800 dimensions de la ressource utilisée. La désambiguïssation d'un terme résulte en une combinaison des vecteurs de ses cooccurrences. Les expériences menées sur la campagne d'évaluation ad-hoc Amaryllis n'ont malheureusement pas montré d'apport, lorsque la désambiguïssation est faite sur les vecteurs sémantiques de documents. Toutefois, l'approche initiale, consistant à combiner les scores de la recherche sémantique avec ceux de la recherche lexicale, reste meilleure qu'une recherche ne faisant pas intervenir la sémantique.

L'utilisation du moteur de recherche "sémantique" dans le cadre d'EquER, l'évaluation des systèmes de questions-réponses en français, s'avère être largement bénéfique. Les documents retournés en tête de liste ont plus de chances de contenir une réponse correcte, ce qui augmente les chances de bien répondre aux questions posées. L'analyse fine des questions autorise à la fois de poser des requêtes plus pertinentes et de générer des reformulations (patrons d'extraction) susceptible de coller un contexte des réponses. L'utilisation des ressources linguistiques dans le calcul de similarité entre questions et contextes des réponses ne permet pas un gain significatif de la précision. Les expériences menées sur le nombre de documents retournés par le moteur de recherche, ont montré que le moteur sémantique joue un rôle important dans le processus de recherche des réponses, car plus le nombre de "documents pertinents" est important en tête de liste, meilleur sera la précision du moteur de questions-réponses.