

AVIS DE SOUTENANCE DE THÈSE DE DOCTORAT

Monsieur Laurent GILLARD soutiendra une thèse
Le jeudi 25 octobre 2007 à 13h 30

L.I.A.

SPÉCIALITÉ : INFORMATIQUE

Titre de la thèse : Quelles méthodes pour les sujets de Questions-Réponses ? Une avancée vers le tout numérique

Membres du jury :

Brigitte GRAU, PR, Informatique, Ecole Nationale Supérieure d'Informatique pour l'Industrie et l'Entreprise, Evry

Pierre ZWEIGENBAUM, DR, Informatique, LIMSI, Université Paris XI

Jacques SAVOY, PR, Informatique, Institut d'Informatique de Neuchâtel, Suisse

Patrice BELLOT, MCF, Informatique, Laboratoire Informatique d'Avignon, Université d'Avignon et des Pays de Vaucluse

Marc EL-BEZE, PR, Informatique, Laboratoire Informatique d'Avignon, Université d'Avignon et des Pays de Vaucluse

Résumé de la thèse :

Les différents travaux décrits dans ce mémoire ont trait aux systèmes de Questions/Réponses, c'est-à-dire à des systèmes de recherche d'information capables de proposer une réponse précise et concise à une question formulée en langue naturelle.

Après avoir introduit les principes des systèmes de Questions/Réponses, l'étude de ces systèmes est envisagée dans le cadre des campagnes d'évaluation, et notamment au travers d'une participation à la campagne francophone EQueR (sur des données en français).

Nous proposons ensuite d'étudier différentes approches numériques pour la sélection de passages et l'extraction de réponses exactes (sous la contrainte de disposer de peu de connaissances a priori) Nous définissons un score pour chacune de ces deux étapes et nous étudions également leur contexte d'utilisation. Le premier score, appelé score de densité, permet une sélection des passages dans les documents ; le suivant, le score de compacité, permet de préférer une réponse compatible avec un type attendu en fonction de sa proximité et de la densité des mots de la question dans un passage. Ces deux scores sont discutés et comparés à d'autres approches provenant de la « Recherche d'Information ». Nous montrons comment étendre le score de compacité à une version probabiliste, avec des performances comparables, et ouvrons des perspectives sur l'expressivité sémantique qu'il est possible de modéliser grâce à cette extension probabiliste. Enfin, nous envisageons des critères permettant une combinaison optimale des scores provenant de la sélection des passages et de l'extraction d'une réponse.

Concernant les questions booléennes, nous esquissons une méthode particulière de résolution combinant quelques connaissances linguistiques et des calculs numériques de parité et polarité.