

Titre : *Sémantique légère pour l'exploration de documents- Assistance personnalisée pour l'accès du document et l'exploration de son contenu.*

Résumé : Cette thèse met en place les principes d'une *sémantique légère* pour le Traitement Automatique des Langues qui suppose une limitation à la fois des ressources et des processus utilisés pour proposer des services personnalisés aux utilisateurs. Elle s'inscrit en cela dans le courant des recherches récentes qui tentent de pallier les inconvénients des objectifs maximalistes de l'IA classique pour la compréhension et le calcul du sens des textes. Les ressources utilisées sont des entités lexicales catégorisées et décrites de façon componentielle et différentielle par l'utilisateur. Ces données sont exploitées pour des calculs fondés sur la récurrence et la différence. Le modèle proposé permet l'expression de besoins spécifiques à un usager ou à un groupe d'utilisateurs dans le cadre de tâches génériques nécessitant l'exploration du contenu des textes. Les ressources du système sont construites en interaction avec la machine ; elles peuvent être acquises à partir d'un corpus. Nos efforts ont porté à la fois sur la modélisation lexicale et sur les moyens de l'interaction entre l'utilisateur et la machine.

Les champs d'évaluation sont la veille documentaire et l'analyse d'un fait de langue (une métaphore conventionnelle).

Cette thèse d'informatique participe aux recherches qui ont pour but de transformer la machine en un média à valeur ajoutée pour l'accès personnalisé aux documents textuels. Elle se situe à l'intersection de trois courants de recherche : celui de la mise en oeuvre de la sémantique lexicale de tradition saussurienne (la Sémantique Interprétative de F.Rastier), celui de l'exploitation assistée de corpus numériques et enfin, celui de l'interaction homme / machine située.

Mots clés : traitement du langage naturel, sémantique, signification, système homme-machine, interfaces utilisateur (informatique).

Title : *Light semantics for document content analysis – Customized services for document content analysis.*

Abstract : This thesis investigates and defends a "light semantics" approach to Natural Language Processing (NLP). The contrast between "light semantics" and the use of large databases and complex processes is described. Our system provides users with customized services based on both limited-quantity resources and simple processes. This study is thus related to the recent works trying to mitigate the disadvantages of the maximum objectives of the classic AI approaches for automatic text understanding. The system resources are lexical entities, which are categorized and described both componentially and differentially. Text processing is then based on recurrences and differences in textual documents. The model allows users to express their specific needs within the framework of generic tasks requiring document content analysis. The system entries are built through interactions between user and computer. They can be acquired from a corpus. Our work has focused both on the lexical model and on the needed interaction between user and computer.

The chosen evaluation fields are document scanning and linguistic fact analysis (a conventional conceptual metaphor).

This work is part of researches aiming at transforming computers into an added value media for customized access to texts and for assistance to text content analysis. It is at the junction of three fields of research : traditional structural lexical semantics of Saussure (Interpretative Semantics of F. Rastier), assisted corpora analysis and situated human / computer interaction.

Key words : Natural Language Processing, semantics, human-machine systems, user interfaces