

Rapport sur la proposition de thèse de Monsieur Philippe Martin : "Exploitation de graphes conceptuels et de documents structurés et hypertextes pour l'acquisition de connaissances et la recherche d'informations".

Utiliser des techniques de représentation de connaissances en recherche d'informations est une idée assez ancienne, mais il y a peu de réalisations convaincantes. Utiliser des techniques concernant la structuration et l'organisation de documents en acquisition de connaissances n'a été abordé, pour ce que j'en connais, que de manière marginale. La réalisation de P. Martin, CGKAT, qui intègre de nombreux logiciels (WORDNET, Thot, CoGiTO, le shell), et il aurait été effectivement irréaliste de tout vouloir réaliser à partir de rien, est un système combinant ces deux approches, et ceci d'une manière convaincante. Le travail de thèse de P. Martin concerne donc des domaines particulièrement actifs et importants de l'informatique : la recherche d'informations, l'acquisition et la représentation de connaissances.

Associer des connaissances à des documents, que ce soit pour les indexer afin de pouvoir faire de la recherche par le contenu, pour pouvoir naviguer dans un ensemble de documents, ou encore pour créer des documents virtuels par assemblage de documents existants, nécessite d'avoir des connaissances pour décrire ces documents et leurs relations. Ainsi l'utilisation de techniques d'intelligence artificielle dans des activités de gestion de documents nécessite une étape d'acquisition et de représentation de connaissances, et en premier lieu la définition d'un ensemble de termes permettant de les décrire : d'une "ontologie" (au sens IA du terme). D'autre part, dans la construction de bases de connaissances il est nécessaire de définir un langage, dont le vocabulaire est constitué par une ontologie. CGKAT propose donc un ensemble de concepts et de relations généraux (les niveaux les plus abstraits d'une ontologie), issu, principalement, de WordNet et de Sowa. CGKAT propose aussi diverses fonctionnalités de gestion d'une ontologie, devant permettre à un utilisateur de construire l'ontologie dont il a besoin, car il est peu probable que l'ontologie fournie le satisfasse pleinement.

Le modèle de représentation de connaissances utilisé par P. Martin est celui des graphes conceptuels, modèle introduit par Sowa en 1984. Je pense que c'est un modèle particulièrement bien adapté aux objectifs de P. Martin. En effet, l'indexation de documents par des graphes conceptuels est une technique qui a déjà été utilisée par plusieurs chercheurs, et P. Martin propose des outils d'aide à la construction de telles indexations, en particulier en utilisant d'une manière astucieuse l'éditeur de documents structurés Thot, ainsi que les diverses fonctionnalités offertes par la plateforme de manipulation de graphes conceptuels, CoGiTO, développée au LIRMM. Ceci lui permet, grâce aux extensions du modèle qu'il a réalisées, de proposer un langage de requêtes "conceptuelles" intéressant.

Deux regrets cependant concernant ce travail. L'un concerne la rédaction de cette thèse : pour réaliser ce travail P. Martin a fait de très nombreuses lectures, et a étudié de nombreux logiciels complexes, malheureusement il a voulu "tout" raconter dans sa thèse. Le résultat est un document touffu, avec des redites (les chapitres sont pratiquement autonomes), et les raisons de certains développements sont parfois difficiles à comprendre. En particulier, la première partie, celle sur l'état de l'art, est, à mon avis trop longue, et risque de susciter des critiques de divers spécialistes sur des points de détail. Par contre, la seconde partie est beaucoup plus agréable à lire, et certains chapitres prouvent que P. Martin sait rédiger. L'absence d'expérimentation, du moins rapportée dans le mémoire, est mon deuxième regret : il est souvent question, dans ce mémoire, d'un cogniticien que CGKAT est censé aider dans diverses activités : quel cogniticien a utilisé CGKAT ? pour faire quoi ? quelles étaient ses critiques et remarques ? ... en bref une validation expérimentale serait nécessaire.

Je pense qu'il faut plutôt incriminer le système actuel des thèses que le candidat : il est pratiquement impossible d'avoir des idées originales, de les implémenter dans un système complexe intégrant des logiciels existants, de rédiger et documenter l'ensemble, et de faire une validation expérimentale, dans le temps imparti aujourd'hui pour une thèse.

En conclusion, par sa réalisation importante, P. Martin a prouvé qu'il était possible, aujourd'hui, de concevoir et réaliser des systèmes combinant des techniques d'intelligence artificielle et de manipulation de documents, offrant de nombreuses fonctionnalités utiles aussi bien en acquisition de connaissances qu'en recherche d'informations, je suis donc entièrement favorable à ce que son travail fasse l'objet d'une thèse en informatique.



Michel Chein

Professeur à l'Université de Montpellier