

Rapport établi par François RECHENMANN, Directeur de Recherche à l'Inria Rhône-Alpes, sur le mémoire intitulé

« Exploitation de graphes conceptuels et de documents structurés et hypertextes pour l'acquisition de connaissances et la recherche d'informations »

que présente Philippe MARTIN

en vue de l'obtention du grade de docteur de l'Université de Nice Sophia-Antipolis

La construction d'une base de connaissances peut être vue comme un processus de modélisation d'un domaine identifié, conformément à des objectifs bien énoncés. Les sources de connaissances sont souvent textuelles, que ces textes aient été spécifiquement écrits lors du processus d'acquisition auprès de spécialistes du domaine ou qu'ils soient extraits de la littérature spécialisée du domaine. Dans les deux cas, il est extrêmement utile de conserver des liens entre ces sources textuelles et les connaissances formalisées dans la base afin en particulier d'aider un utilisateur à en comprendre le contenu ou d'expliquer les résultats obtenus par un mécanisme de résolution de problèmes qui exploiterait les connaissances qu'elle contient. Le mémoire que présente Philippe Martin décrit précisément un environnement informatique, CGKAT, qui permet de créer et de maintenir des liens entre des connaissances formalisées à l'aide de graphes conceptuels et des documents structurés.

Ce mémoire est composé de deux grandes parties. La première présente en trois chapitres l'état de l'art en matière d'acquisition des connaissances, de structuration et de recherche d'informations et de graphes conceptuels. Cet état de l'art repose sur des études bibliographiques sérieuses et importantes comme en atteste la liste des références qui figure à la fin du mémoire.

Les contributions de Philippe Martin sont l'objet de la seconde partie. Il s'agit d'une part du développement d'une ontologie générale et réutilisable, intégré dans l'environnement CGKAT, d'autre part de la description de l'environnement proprement dit et de son implémentation. L'existence d'un tel environnement opérationnel est un apport important au domaine de l'acquisition et de la représentation des connaissances.

Dans son chapitre présentant l'état de l'art en matière d'acquisition des connaissances, Philippe Martin rappelle qu'une ontologie rassemble les termes utilisables pour représenter les connaissances d'un domaine. Un des thèmes de recherche actuels est la construction d'ontologies réutilisables. Notons que ce principe de réutilisabilité peut être contesté, dans la mesure où une base de connaissances est considérée comme un modèle, dont l'élaboration est extrêmement dépendante des objectifs fixés, autrement dit d'un ensemble de questions spécifiques auxquelles la construction de la base doit tenter de répondre. Philippe Martin présente donc une ontologie générale, destinée à faciliter la construction de bases de connaissances dans des domaines divers et non précisés *a priori*. Pour ce faire, il s'est inspiré de travaux similaires et propose une connexion à la base de connaissances terminologiques WordNet. Au delà de la question sur le principe même de réutilisabilité, il est évidemment difficile d'évaluer la pertinence d'une telle ontologie.

L'environnement CGKAT repose sur Thot (développé au sein du projet Opéra de l'Inria Rhône-Alpes) pour l'édition et la manipulation de documents structurés et sur CoGITO (développé au LIRMM à Montpellier) pour la manipulation de graphes conceptuels. À travers cette réalisation importante, Philippe Martin met bien en évidence bien tous les avantages à disposer, au sein d'un même environnement, des fonctionnalités d'édition de textes et de représentation formelle de connaissances. En particulier, les connaissances formalisées par des graphes conceptuels peuvent être considérées comme indexant des documents. Il est ainsi possible de diriger la recherche d'informations par des requêtes à la base de connaissances. L'existence d'un tel

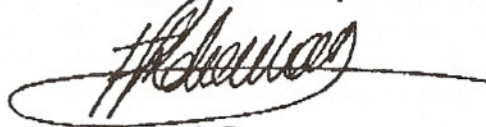
environnement opérationnel constitue donc une contribution majeure et bien identifiable au domaine de l'acquisition des connaissances.

Enfin, la conclusion du mémoire reprend les différentes contributions et les situe dans le paysage de l'existant. Elle est d'autant plus appréciée que le mémoire est long, au point de diluer ces contributions dans les considérations générales de la première partie et les détails techniques de la seconde.

Des annexes d'une centaine de pages complètent ce mémoire. Il est surprenant d'y trouver mélangés, d'une part des détails sur les ontologies discutées dans le corps du mémoire ou des descriptions d'interfaces fonctionnelles, et d'autre part quelques pages sur la construction coopérative et incrémentale de hiérarchies de types, pages qui mériteraient sans aucun doute d'être développées et de figurer ensuite en bonne place dans le mémoire.

Le mémoire de Philippe Martin décrit des travaux d'envergure qui touchent simultanément trois domaines : l'acquisition des connaissances, leur formalisation à l'aide des graphes conceptuels et la recherche d'informations. Ces travaux se situent au cœur de la problématique actuelle de la construction de grandes bases de connaissances. Ils se concrétisent par une réalisation informatique importante et pertinente, support permettant l'étude des liens entre connaissances textuelles et formelles. Pour ces raisons, je donne un avis favorable à ce que ces travaux fassent l'objet d'une soutenance en vue de l'obtention du titre de docteur en informatique de l'Université de Nice Sophia-Antipolis.

Grenoble, le 30 septembre 1996



François REICHENMANN