

BONOM : un système Multi-Agents pour la recherche d'informations sur Internet dirigée par la connaissance.

Sylvie Cazalens, Emmanuel Desmontils, Christine Jacquin et Philippe Lamarre
IRIN, Université de Nantes, 2 rue de la Houssinière BP92208, 44322 Nantes Cedex 3,
<http://www.sciences.univ-nantes.fr>, {cazalens,desmontils,jacquin,lamarre}@irin.univ-nantes.fr

L'objectif général du projet BONOM [1,2] est d'améliorer la gestion, la recherche et la consultation d'information sur le réseau Internet, sans toutefois le surcharger. BONOM est une *société d'agents d'information intelligents et collaboratifs* organisée en *système fédéré* : c'est une structure hiérarchique où des agents dits « *facilitateurs* » ou « *médiateurs* » dirigent les requêtes selon leur nature depuis les *agents personnels* (l'utilisateur) vers les *agents sites* (les sites web). Cependant, à la différence des systèmes fédérés classiques, chaque fédération (que nous appelons « communauté simple ») possède plusieurs facilitateurs dont les rôles sont identiques. De façon à diriger les requêtes de recherche d'informations le plus rapidement et le plus pertinemment vers les sites concernés, la société est donc organisée en différentes *communautés simples*, éventuellement imbriquées, qui sont spécialistes d'un domaine informationnel particulier appelé *thème*. Les agents d'une communauté simple ne traiteront que les requêtes qui concernent leur domaine, les autres seront proposées aux communautés imbriquées. La notion qui relie ici les agents d'une même communauté simple est donc celle de domaine informationnel. Dans un premier temps, nous avons choisi de considérer une structure de domaines particulière qui ressemble à celles des « newsgroups » actuels. Un domaine informationnel sera donc une succession de thèmes, chacun indiquant une spécialisation du thème qui précède. Notons que nous ne nous intéressons pas pour l'instant aux sociétés auto-organisatrices qui décideraient elles-mêmes s'il convient de rajouter un domaine informationnel. Nous considérons que comme dans les *newsgroups*, ce sont les utilisateurs et les responsables de la société qui décident. Cependant, nous nous proposons de fournir des outils de haut niveau pour aider les intervenants humains dans la gestion de la société (organisation des agents, gestion de la connaissance...).

Toute la connaissance dans Bonom est représentée par des ontologies. Elle est divisée en deux types: la connaissance associée à la structuration des domaines (les domaines sont les concepts) et la connaissance propre aux domaines sous forme d'un ensemble d'ontologies. Dans la société Bonom, la connaissance est disponible via deux types d'agent fournisseurs: les agents fournisseurs d'ontologie et les agents fournisseurs de domaine. Afin de déterminer le ou les domaines potentiels de son site, le gestionnaire de site doit avec l'aide des agents fournisseurs de domaine et d'ontologie définir l'ensemble des ontologies qui vont servir à indexer son site. Dans ce but, le gestionnaire définit des domaines, à partir desquels les agents fournisseurs proposent des ensembles d'ontologies leur correspondant. Les ontologies sont pondérées à l'aide de coefficients qui mesurent leur corrélation avec certains domaines. Ce coefficient permet au gestionnaire de sélectionner des ontologies qui seront utilisés par l'agent indexeur pour indexer le site.

Le processus d'indexation consiste à extraire les concepts des pages web et à les associer avec les pages dans lesquelles ils apparaissent ainsi qu'avec certains concepts des ontologies sélectionnées précédemment. Le processus d'indexation extrait d'abord les termes des pages en utilisant des techniques de traitement automatique des langues (étiquetage, repérage des syntagmes nominaux) et les informations inhérentes au marquage HTML. Ensuite, à partir des termes précédemment acquis, les concepts correspondants sont déterminés en utilisant un thesaurus (Wordnet). Un terme est associé avec un concept du thesaurus si ce terme fait parti du synset associé au concept et si des termes qui sont dans la même page, sont en relation conceptuelle avec le terme étudié. Finalement, parmi tous les concepts potentiels, le(s) plus pertinent(s) est (sont) retenu(s). Le thesaurus est utilisé pour réaliser l'appariement entre les concepts de l'ontologie et les concepts déterminés dans les pages. Pour cette raison, tout concept doit avoir un attribut qui est une de ses représentations lexicales et qui va permettre de déterminer dans le thesaurus le concept correspondant (le synset). Il y aura appariement si le concept de l'ontologie apparaît dans les pages ainsi que d'autres concepts qui lui sont reliés. Si à la fin du processus d'appariement certains concepts extraits des pages n'ont pu s'apparier avec un concept d'une ontologie, le gestionnaire de site peut redemander aux agents fournisseurs de domaines et d'ontologie de nouvelles ontologies relatives à ce concept. Un certain nombre de mesures permettent au gestionnaire de site de sélectionner les ontologies les plus représentatives du contenu de son site ainsi que les domaines pertinents.

Références

- [1] S. Cazalens, E. Desmontils, C. Jacquin & P. Lamarre, « Structured Distributed Middle Agents for Information Retrieval on the Internet », RR-IRIN-00.2, IRIN, Janvier 2000.
- [2] S. Cazalens, E. Desmontils, C. Jacquin & P. Lamarre, « A Distributed Agent-Based System for Knowledge-Based Internet Information Retrieval », Internet Computing'2000, Las Vegas, USA, 26-29 juin 2000, à paraître