
Accès personnalisé multicritères à de multiples sources d'informations.

Samir kechid

*Université des Sciences et de la Technologie Houari Boumediene. USTHB BP 32 El Alia Bab Ezzouar Alger Algérie.
kechidsam@yahoo.fr*

RÉSUMÉ. Cet article décrit une approche de la recherche d'information permettant l'accès personnalisé à plusieurs sources d'informations. L'accès à des sources d'informations distribuées est souvent effectué en trois étapes, la première consiste à sélectionner les sources pertinentes pour la requête, puis soumettre la requête à ces sources sélectionnées et finalement fusionner les résultats retournés par ces sources. L'objectif de cet article est d'intégrer l'utilisateur via son profil dans les processus de sélection et de fusion des résultats des sources.

ABSTRACT. This article describes an approach for information retrieval giving access personalized to several sources of information. The access to distributed sources of information is often carried out in three stages, the first consists in selecting the relevant sources for the query, then to subject the query to these selected sources and finally to merging the results turned over by these sources. The objective of this article is to integrate the user via his profile in the sources selecting and results merging process.

MOTS-CLÉS: recherche information distribuée, agents intelligents, profil source, profil utilisateur, sélection de sources, fusion de résultats.

KEYWORDS: distributed information retrieval, source profile, user profile, server selecting, result merging.

1. Introduction

Avec le développement des réseaux de communications, nous observons ces dernières années une multiplication des sources (serveurs) d'information hétérogènes. L'accès à ces sources inclus trois sous problèmes : 1) acquisition des informations sur le contenu de la source « description de la source » (Callan 2000) (Nottelmann *et Fuhr* 2003) (Si *et Callan* 2005), 2) sélection d'un nombre minimum des sources pour une requête donnée (sélection de sources) (Xu *et al* 1998) (Callan 2000) (Gravano *et al* 1999) (Nottelmann *et Fuhr* 2003) (Si *et Callan* 2005), et 3) fusion des résultats des sources et affichage d'une seule liste de résultats à l'utilisateur « fusion de résultats » (Lark *et al* 2000) (Xu *et al* 1998) (Si *et Callan* 2003). Ces trois sous problèmes sont souvent résolus, prenant en compte uniquement la requête de l'utilisateur. L'utilisateur qui a exprimé ce besoin en information est mis à l'écart du processus.

L'objectif de notre travail est de définir une approche permettant d'intégrer l'utilisateur via son profil dans les processus de sélection et de fusion des résultats des sources.

Cet article est destiné à la conférence des jeunes chercheurs, il ne présente pas des résultats finaux, il présente notre axe de recherche et les objectifs que nous voulons atteindre par notre travail de thèse.

2. Notre approche

Notre contribution consiste à définir d'un côté un profil utilisateur pour décrire le centre d'intérêt et les préférences de l'utilisateur, de l'autre côté de définir un profil de la source pour décrire la source. Le profil utilisateur et le profil de la source sont utilisés dans le processus de sélection de source et de fusion des résultats, pour délivrer à l'utilisateur des résultats pertinents à sa requête tout en satisfaisant au maximum son centre d'intérêt et ses préférences.

Nous présentons notre démarche par les tâches suivantes :

- 1- Définir un profil pour chaque utilisateur, qui permet de décrire le centre d'intérêt de l'utilisateur et ses préférences
- 2- Définir un profil pour chaque source d'information, qui permet de décrire la source.
- 3- Définir une méthode pour apparier le profil utilisateur avec les profils des sources, dans le but de sélectionner pour cet utilisateur les sources pertinentes.
- 4- Interroger les sources sélectionnées avec la requête utilisateur.
- 5- Définir une méthode pour fusionner et classer les résultats retournés par les sources, en prenant en compte le profil utilisateur.

Nous présentons l'architecture générale de notre approche par la figure suivante (N_s est le nombre de sources):

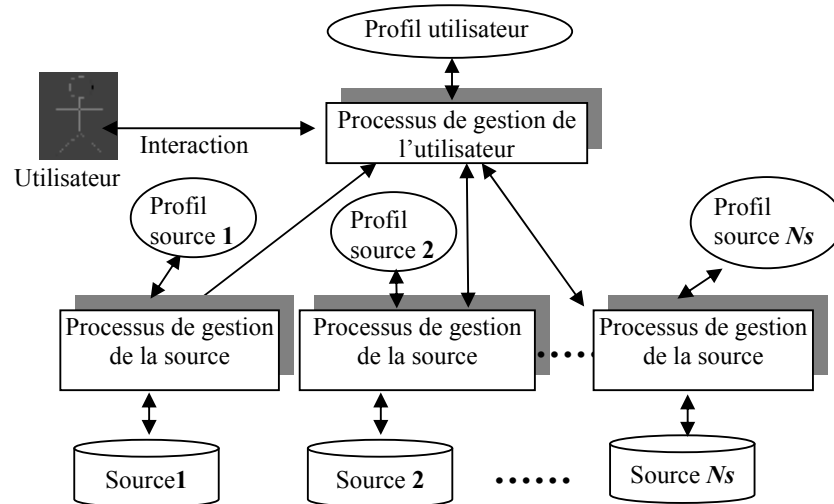


Figure 1 : architecture générale

L'architecture de la figure 1 décrit le fonctionnement global de notre approche. Le processus de gestion de l'utilisateur assure les tâches 1, 3 et 5, en interaction avec le processus de gestion de la source. Le processus de gestion de la source assure les tâches 2 et 4.

Pour le fonctionnement de notre approche, nous nous sommes orienté vers les agents intelligents. Ce choix a été fait sur la base des caractéristiques qu'ils possèdent et qui s'adaptent mieux à notre approche. Nous pouvons résumer les caractéristiques d'un agent intelligent comme suit :

- *L'autonomie*
- *Capacité à communiquer et à coopérer.*
- *Capacité à raisonner, à réagir à leur environnement.*

Plusieurs approches ont été développées sur la base des agents intelligents pour l'apprentissage de l'utilisateur (Baldwin *et al* 2000) (Chen *et Sycara* 1998) (Seo *et Zhang* 2000). D'autres approches ont utilisées les agents intelligents pour capturer les informations distribuées (Ogston 2003) (Yu *et Singh* 2003) (Zhang *et al* 2004).

Le rôle des agents intelligents dans notre approche est de capturer les informations de l'utilisateur et de son environnement pour construire son profil, de capturer les informations des sources pour construire le profil de chaque source, de décider quelles sont les sources à utiliser pour l'utilisateur lors de ses recherches en information, et lui délivrer des informations pertinentes.

Nous distinguons suivant nos besoins deux types d'agents,

1- Des agents sources. Un agent source s'occupe de la source, en fournissant à l'agent utilisateur les informations de la source dont il a besoin pour accomplir ses tâches. Cet agent assure le fonctionnement du processus de gestion de source. Un agent est associé à une source.

2- Des agents utilisateurs. Un agent utilisateur s'occupe de l'utilisateur, en répondant à ses besoins en informations. Cet agent assure le fonctionnement du processus de gestion de l'utilisateur. Un agent utilisateur est associé à un utilisateur.

Nous présentons dans ce qui suit la description générale du profil utilisateur et du profil source.

2.1. Description du profil de l'utilisateur

Plusieurs approches ont été développées pour définir le profil utilisateur, nous pouvons citer : les approches adaptatives (Rochio 1971) WebMate (Chen *et* Sycara 1998) (Poo *et al* 2003) ; les approches sémantiques (Billsus *et* Pazzani 1999) (Pretschner *et* Gauch 1999) (Gauch *et al* 2003) ; approche multidimensionnelle (kostadinov 2003)

Pour mieux exploiter le profil utilisateur tout en l'adaptant à nos besoins nous le divisons en trois dimensions. La première dimension représente *l'historique des interactions de l'utilisateur*, la deuxième représente le *centre d'intérêt* de l'utilisateur et la troisième représente *les préférences de l'utilisateur*.

Nous avons défini la représentation globale que nous adoptons pour décrire un utilisateur. La suite de notre travail consiste à :

- 1- Comment instancier ces trois dimensions du profil ?
- 2- Comment exploiter ce profil dans le processus de sélection et de fusion.

2.2. Description du profil source

Le profil source doit contenir un maximum d'informations pour décrire la source. Nous avons proposé de définir le profil source en deux dimensions principales. Une dimension pour décrire le contenu de la source, nommée *Thèmes*, l'autre pour décrire les caractéristiques de la source nommée *Critères*.

- *Thèmes* : les thèmes décrivent le contenu de la source.
- *Les critères* : les critères décrivent les caractéristiques de la source.

Nous avons défini la représentation globale que nous adoptons pour décrire une source d'information. La suite de notre travail consiste à répondre à ces questions :

- 1- Que doit contenir la dimension «*Thèmes*» ?

2- Que doit contenir la dimension «*Critères*» ?

3- Comment «appairer» entre le profil utilisateur et le profil des sources, pour choisir les sources pertinentes à un utilisateur ?

3. Conclusion

Cet article décrit les objectifs que nous voulons atteindre par notre travail de thèse. Qui se caractérise à définir une approche d'accès personnalisée dans un environnement multi sources. Pour ce faire nous nous sommes orientés vers la définition d'un profil de la source d'un côté, de l'autre côté la définition d'un profil utilisateur, puis de définir une méthode pour «appairer» le profil utilisateur avec les profils des sources pour délivrer à l'utilisateur les informations pertinentes suivant son centre d'intérêt et ses préférences personnelles.

Pour le fonctionnement de notre approche, nous avons proposé d'utiliser des agents intelligents, un agent par source et un agent par utilisateur. L'agent de chaque source a pour rôle de maintenir à jour le profil de la source, et de fournir à l'agent utilisateur les informations nécessaires pour accomplir ses tâches. L'agent utilisateur a pour rôle de maintenir à jour le profil de l'utilisateur, lui trouver les sources d'informations pertinentes et lui délivrer les informations pertinentes.

Nous avons défini une représentation globale d'un profil source, et d'un profil utilisateur, notre travail consiste par la suite à décrire le contenu de chaque profil, et de définir une méthode pour «appairer» le profil utilisateur avec le profil des sources, afin de trouver les sources et les informations qui sont pertinentes pour l'utilisateur.

Bibliographie

- Billsus D, Pazzani J. A Hybrid User Model for News Story Classification. 7-th International Conference on User Modeling (UM 99), Banff, Canada, June 20-24, 1999.
- Callan J. Distributed information retrieval. In W.B. Croft, editor, *Advances in Information Retrieval*. Kluwer Academic Publishers. (pp. 127-150), 2000.
- Chen L, Sycara K. WebMate: A Personal Agent for Browsing and Searching, In *Proceedings of the 2nd International Conference on Autonomous Agents and Multi Agent Systems*, Minneapolis, MN, May 10-13, 1998.
- Gauch S, Chaffé J, Pretschner A. Ontology-Based User Profiles for Search and Browsing, To appear in *J. User Modeling and User-Adapted Interaction, the Journal of Personalization Research*, Special Issue on User Modeling for Web and Hypermedia Information Retrieval, 2003.
- Gravano L, Garcia-Molina H, Tomasic A. GLOSS: Text-source discovery over the Internet, *ACM Transactions on Database Systems*, 24(2), (pp. 229-264, 1999).

- Kostadinov D. Personnalisation de l'information et gestion des profils utilisateurs. Mémoire de DEA PRiSM , Versailles. 2003.
- Larkey L.S, Connell M.E, Callan J. Collection and Results Merging with Topically Organized U.S. Patents and TREC Data. Proceedings of CIKM'2000. (pp.282-289), 2000.
- Nottelmann H et Fuhr N. Evaluating different method of estimating retrieval quality for resource selection. In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information retrieval, 2003.
- Ogston. E. Group formation among peer-to-peer agents: learning group. Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, 2003.
- Poo D, Chang B, Goh J. A Hybrid Approach for User Profiling. 36th Annual Hawaii International Conference on System Sciences (HICSS'03) - Track 4, January 06 - 09, 2003.
- Pretschner A, Gauch S. Ontology Based Personalized Search. In Proceedings of the 11th IEEE International Conference on Tools with Artificial Intelligence (ICTAI), November 1999.
- Rochio J. Relevance feedback information retrieval. In Gerald Salton (editor), The SMART retrieval system- experiments in automated document processing. Prentice-Hall, Englewood Cliffs, NJ, 1971.
- Seo Y-W, Zhang T A Reinforcement Learning Agent for Personalized Information Filtering, Proceedings of the 2000 International Conference on Intelligent User Interfaces, New-Orleans, USA, Jan 9-12, 2000, ACM, pp248-251, 2000.
- Si L and Callan J. A Semi-Supervised learning method to merge search engine results. ACM Transactions on Information Systems, 21(4). (pp. 457-491), 2003.
- Si L and Callan J. Modeling Search Engine Effectiveness for Federated Search. In Proceedings of the Twenty Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Toronto, Canada: ACM 2005.
- Xu J, Callan J.P. Effective retrieval with distributed collections. ACM SIGIR'98, (pp. 112-120), 1998.
- Yu. B and Singh. M. Incentive mechanisms for agent-based peer-to-peer systems. Proceedings of the Second International Joint Conference on Autonomous Agents and Multiagent Systems, 2003.
- Zhang H, Croft W.B, Levine B, Victor Lesser. A Multi-agent Approach for Peer-to-Peer-based Information Retrieval Systems AAMAS'04, July 19-23, 2004, New York, New York, USA. Copyright 2004 ACM 1-58113-864-4/04/0007. 2004.