
Recherche contextuelle d'information

Mariam Daoud

*IRIT- Equipe SIG/RI
118 route de Narbonne
31062 Toulouse cedex 4
daoud@irit.fr*

RÉSUMÉ. La recherche contextuelle d'information a pour objectif de mieux répondre aux besoins de l'utilisateur et de lui délivrer de l'information pertinente selon son contexte spécifique de recherche. Ce papier présente un contexte de recherche à deux dimensions : le contexte de l'utilisateur et le contexte de la requête. On vise à intégrer ces deux dimensions du contexte dans un processus de recherche d'information afin d'asseoir un système de recherche d'information contextuelle.

ABSTRACT. Contextual retrieval information aims to better meet the user's needs and to deliver relevant information to him according to his specific retrieval context. This paper presents a retrieval context in two dimensions: user context and query context. We aim to integrate these two dimensions of context into a process of information retrieval to sit a contextual information retrieval system.

MOTS-CLÉS: recherche d'information, contexte de l'utilisateur, contexte de la requête.

KEYWORDS: information retrieval, query context, user context.

1. Introduction

La popularisation d'Internet ainsi que l'explosion des services de l'information de nos jours ont propulsé la recherche d'information (RI) au premier plan. En effet, la surabondance de l'information a engendré la dégradation de la qualité des résultats retournés par un système de recherche d'information (SRI) et a apporté des nouveaux problèmes au domaine de la RI. (Budzik et al., 00) montre que l'approche généraliste non « personnalisée » des outils disponibles de RI sur le Web est à l'origine des problèmes évoqués, où seule la requête représente complètement l'utilisateur. C'est pourquoi des travaux sur la RI contextuelle ont vu le jour ces dernières années dont le but fondamental est de combiner les technologies de recherche et les connaissances portant sur la requête et sur le contexte de l'utilisateur dans une même infrastructure afin de retourner une réponse appropriée à ses besoins (Allan, 02). Les chercheurs dans le domaine de l'adaptation au contexte en recherche d'information n'ont pas encore abouti à une définition complète et générique de la notion du contexte et plus précisément les éléments qui le constituent. Certains travaux (Lawrence, 00) et (Ingwerson et al., 04) portent sur la définition du contexte comme l'ensemble de facteurs cognitifs, sociaux et des facteurs reliés aux tâches, aux intentions et aux buts de l'utilisateur durant les épisodes de recherche d'information. Or la formalisation du contexte et son exploitation dans un processus de RI diffère entre les travaux de RI contextuelle. (Huang, 05) développe une infrastructure extensible de RI contextuelle dans laquelle le contexte est incorporé au modèle probabiliste okapi BM25. D'autres travaux en RI contextuelle modélisent l'utilisateur selon un profil traduisant ses centres d'intérêts à court et long terme et ses préférences et utilise ce profil dans au moins une phase du processus de RI (Bottraud et al. 03, Rozé et al. 00). Suite à l'effort imposé à l'utilisateur lors de la saisie explicite de ses préférences, des travaux en RI contextuelle ont recours à la prédiction des préférences des résultats retournés par réinjection de pertinence implicite (Agichtein et al., 06). Ce dernier travail utilise des données de clics et de navigation pour prédire l'ordre de préférences des résultats.

Notre objectif est de proposer des techniques permettant de supporter un processus de RI contextuelle. Plus précisément, nous envisageons d'exploiter d'autres sources d'évidence que la requête afin de mieux cerner les besoins spécifiques d'un utilisateur donné. Pour cela on vise à construire un « contexte » de recherche qui consiste en un ensemble d'informations additionnelles et l'intégrer dans un processus de RI. Ces informations additionnelles sont divisées en deux catégories : la première est liée à l'utilisateur et la deuxième est liée à la requête. On note que des travaux dans le domaine de la clarification des besoins des utilisateurs via des interfaces interactives, ont utilisé des aspects liés à la requête tels que le profil temporel et le profil thématique afin d'améliorer la précision de recherche. Cependant, ces améliorations sont dépendantes de la requête et non de l'utilisateur d'une manière spécifique (Rode et al., 06). En revanche, notre contribution consiste à clarifier la requête en exploitant son contexte en termes de centres d'intérêts et des préférences de l'utilisateur qui l'a émise.

Ce papier est organisé comme suit : la section 2 explore les dimensions de notre contexte. La section 3 montre la phase de construction de ce contexte. La section 4 montre l'exploitation de ce contexte dans un processus de RI. En dernier une conclusion résume notre contribution et présente nos perspectives de recherche.

2. Dimensions du contexte

Dans le cadre de notre travail, on propose de définir deux principales dimensions du contexte de recherche. La première permet de caractériser l'utilisateur à travers ses centres d'intérêts à court et à long terme et ses préférences. La seconde dimension est liée à la requête et permet de cibler la nature des résultats attendus par l'utilisateur.

2.1. Contexte lié à l'utilisateur

On définit le contexte lié à l'utilisateur par des éléments qui peuvent être donnés explicitement par l'utilisateur ou déduits implicitement par le système. On cite les centres d'intérêts qui peuvent appartenir aux domaines de la science, sport, humanité, etc., et les préférences liées à la qualité des informations attendues. Ces préférences sont données explicitement et comportent les deux aspects suivants :

- la fraîcheur : l'utilisateur peut s'intéresser à l'information la plus récente ou bien à des documents liés à des dates bien précises,
- la granularité : l'utilisateur peut s'intéresser à un certain niveau de détail de la réponse attendue, ce peut déterminer la structure de la réponse retournée : un document intégral ou une partie de document (passage), une page d'accueil ou une page cible.

2.2. Contexte lié à la requête

On définit ce contexte par différents paramètres descriptifs et éventuellement mesurables à partir de la requête ou à partir du profil des n top documents retournés par la requête. Soient les paramètres suivants :

- Clarté de la requête : ce paramètre quantifie le degré d'ambiguïté de la requête et permet de mieux cibler la recherche,
- Distribution de la datation des documents : c'est la distribution des n meilleurs documents selon le paramètre date, et qui permet de mettre en évidence la fraîcheur de l'information,
- Degré de couverture de la requête par les services web : ce paramètre permet d'orienter la réponse retournée vers un document ou service.

3. Construction du contexte

La mise en œuvre du processus de RI contextuelle exige la construction du contexte de recherche. Notre contexte sera composé des éléments suivants : le besoin derrière la requête, les centres d'intérêts et les préférences.

3.1. *Le besoin derrière la requête*

On se réfère à l'identification du besoin derrière la requête par une stratégie de classification automatique des requêtes des utilisateurs en trois types :

– *requête informationnelle*: l'intention de l'utilisateur est de trouver de l'information disponible sur le web dans une forme statique et dans plusieurs pages. Aucune interaction n'est prévue que la lecture. On donne comme exemple à ce type de requêtes celles qui demandent des informations dans les domaines de science, médecine, histoire, etc.

– *requête transactionnelle*: le but général est d'accéder à un site où une interaction ultérieure va suivre. Cette interaction constitue la transaction définissant ce type de requête. Les catégories principales de ce type de requête sont achats, trouver des services via le web, télécharger plusieurs types de fichiers (images, chansons, etc.), accéder à des bases de données (pages jaunes), trouver des serveurs (jeux), etc.

– *requête navigationnelle*: l'intention de l'utilisateur est de rechercher un site d'accueil qu'il a dans son esprit ou qu'il a déjà visité dans le passé ou bien en supposant qu'un tel site existe. La plupart des requêtes qui contiennent des noms de compagnies, universités ou des organisations sont considérées navigationnelles.

3.2. *Les centres d'intérêts*

L'observation de l'activité de l'utilisateur permet de dégager implicitement ses centres d'intérêts. On vise à exploiter:

– *l'historique des requêtes*: on considère que les requêtes soumises par l'utilisateur lors des sessions de recherche précédentes permettent de dégager ses centres d'intérêts. Par exemple, si l'utilisateur soumet la requête « java », il est impossible pour le système de savoir si l'utilisateur évoque le langage de programmation « java » ou bien l'île « java » en Indonésie sans aucune information additionnelle. Si l'utilisateur a déjà soumis la requête « CGI programming » durant une session de recherche précédente, le système peut prédire que l'utilisateur recherche les documents qui traitent le langage de programmation java.

– *l'historique d'interactions* : les interactions de l'utilisateur avec le système sont considérées des bonnes sources d'informations permettant d'assigner aux documents visualisés un score de pertinence implicite et par suite d'y dégager les centres

d'intérêts. Ces interactions comportent la consultation, la sauvegarde, l'impression des pages ainsi que des mesures de clics et de navigation (fréquence de clics, temps passé sur une page ou un domaine, etc.).

3.3. Les préférences

On vise à inférer les préférences de l'utilisateur en termes de fraîcheur en exploitant le contexte de la requête plus précisément, le profil temporel de la requête défini comme étant la distribution des n meilleurs documents retournés par la requête selon la dimension temps. L'analyse de ce profil permet de mesurer l'importance du critère « fraîcheur » dans les résultats retournés pour la requête.

4. Exploitation du contexte

Une fois que le contexte est formalisé, il est exploité dans le processus de RI afin d'améliorer la pertinence de l'information retournée. En effet, si le besoin derrière la requête est informationnel, le système applique à celle-ci un test d'ambiguïté en mesurant un score de clarté associé. Ce score mesure la cohérence des résultats retournés par la requête par rapport aux centres d'intérêts de l'utilisateur. En effet le système assume que la requête est ambiguë si ses n meilleurs documents retournés se rapportent à des centres d'intérêts différents. Le système identifie les requêtes ambiguës et les traite différemment, il applique à celles-ci une technique de reformulation en utilisant les centres d'intérêts identifiés à partir des requêtes précédentes. L'exploitation du contexte dans la phase de l'appariement requête-document revient à modifier le calcul du score de pertinence du document en fonction des centres d'intérêts de l'utilisateur. D'autre part, si le besoin en service ou bien les préférences en fraîcheur sont requis par l'utilisateur alors le système prend en considération ces éléments du contexte au niveau du réordonnement des résultats de recherche afin de privilégier les pages transactionnelles ou les documents frais en haut de la liste des résultats retournés.

5. Conclusion et perspectives

Ce papier présente une contribution à la définition d'un contexte de recherche et son exploitation dans un processus de RI. Un aperçu des approches de la littérature en RI contextuelle est tout d'abord abordé. Ensuite, on a défini un contexte à deux dimensions ainsi que les éléments qui le constituent. La construction de ce contexte et son exploitation dans un processus de RI sont explicités, ce qui met en évidence les relations entre ses différents éléments. Nos perspectives de recherche portent sur l'adoption d'une stratégie de classification automatique des requêtes et sur l'utilisation d'une fonction optimisant l'ordre de classement des résultats en fonction

de certains éléments du contexte. En outre, on envisage d'évaluer expérimentalement notre approche selon des protocoles d'évaluations proches de TREC¹.

6. Bibliographie

- Agichtein E., Brill E., Dumais S.T., « Improving Web Search Ranking by Incorporating User Behavior Information », in *the Proceedings of SIGIR*, 2006.
- Allan J. et al., Challenges in information retrieval and language modelling, Report of a Workshop held at the Center for Intelligent Information Retrieval, University of Massachusetts Amherst, september 2002.
- Bottraud J.C., Bisson G., Bruandet M.F., « Apprentissage des profils pour un agent de recherche d'information », *Actes de la Conférence Apprentissage (CAP 2003)*, July 2003, p. 31-46.
- Budzik J., Hammond K.J., « User interactions with every application as context for just-in-time information access », In *Proceedings of the 5th international conference on intelligent user interfaces*, Mars 2000, p. 44-51.
- Gowan J.P Mc., A multiple model approach to personalized information access, Master thesis in computer science, Faculty of science, University college Dublin, 2003.
- Huang X., « Incorporating Contextual Retrieval into Okapi », In *Proceedings of the ACM SIGIR 2005 Workshop on information retrieval in context*, 2005, p. 21-23.
- Ingwersen P., Jarvelin K., « Information retrieval in context », In *Proceedings of the 27th ACM SIGIR Workshop on information retrieval in context*, July 2004, p. 6-8.
- Lawrence S., « Context in web search », *IEEE Data Engineering Bulletin*, vol. 23, n° 3, 2000, p. 25-32.
- Petit-Rozé C., Strugeon E., Abed M., Uster G., Kolski C., « Recherche d'informations personnalisées », *Conférence Internationale NîmesTIC 2000 Ingénierie des Systèmes et NTIC*, Nîmes, France, septembre 2000, p. 401-407.
- Rode H., Hiemstra D., « Using Query Profiles for Clarification », In *Proceedings of the 28th European Conference on Information Retrieval (ECIR)*, 2006.

1. Text Retrieval Conference, disponible sur <http://trec.nist.gov>.