
Recherche contextuelle d'information dans un environnement mobile

Ourdia Boudghaghen

*IRIT-Equipe SIG/RI
118 Route de Narbonne
31062 Toulouse cedex 04
boudgha@irit.fr*

RÉSUMÉ. La recherche contextuelle d'information (RCI) a pour objectif de mieux répondre aux besoins de l'utilisateur en lui délivrant l'information adaptée à son contexte spécifique de recherche. Cet article présente une approche de RCI dans le cas d'un environnement mobile, où le contexte spatio-temporel de l'utilisateur et son contexte cognitif, sont à la fois considérés pour lui délivrer de l'information pertinente. Nous proposons de construire des profils situationnels basés sur l'historique de recherche annoté par le contexte spatio-temporel pour personnaliser les résultats de recherche. De plus, le contexte spatio-temporel sera exploité pour mieux répondre à des requêtes sensibles au temps ou à la localisation.

ABSTRACT. The goal of contextual information retrieval (CIR) is to better meet the user's needs by delivering him information adapted to his specific search context. This paper presents a CIR approach for the case of a mobile environment where the spatio-temporal context of the user and his cognitive context are taken into account to deliver him some relevant information. We suggest building situational profiles based on the research history annotated by the spatio-temporal context to personalize the search results. Furthermore, the spatio-temporal context will be exploited to better address time or location sensitive queries.

MOTS-CLÉS : recherche contextuelle d'information, contexte mobile, profil situationnel, localisation, temps.

KEYWORDS: contextual information retrieval, mobile context, situational profile, location, time.

Ourdia Boudighaghen

1. Introduction

La surabondance de l'information et sa large accessibilité à travers le web, ont engendré la dégradation de la qualité des résultats retournés par les moteurs de recherche d'information (Dervin *et al.*, 86), (Shamber, 94). En clair, le problème n'est pas tant la disponibilité de l'information, mais sa pertinence relativement à un contexte d'utilisation spécifique. C'est pourquoi les travaux en recherche contextuelle d'information (RCI) ont vu le jour ces dernières années (Ingwersen *et al.*, 05), dans le but d'optimiser la pertinence des résultats de recherche, en impliquant deux étapes liées : définition du contexte du besoin en information de l'utilisateur, puis adaptation de la recherche en le prenant en considération dans le processus de sélection de l'information. Diverses taxonomies du contexte ont été proposées dans la littérature pour mieux comprendre les facteurs contextuels qui sont nécessaires pour une meilleure application du contexte dans les systèmes de RI (Göker *et al.*, 08), et (Tamine-Lechani *et al.*, 09) où une taxonomie est définie selon les 5 dimensions spécifiques du contexte qui ont été explorées dans la littérature de la RCI : la *dimension dispositif* (Göker *et al.*, 02), la *dimension tâche/problème* (Jansen *et al.*, 07), la *dimension contexte du document* (Xie, 08), la *dimension spatio-temporelle* (Tao *et al.*, 03), la *dimension contexte utilisateur* qui comprend à son tour deux sous dimensions : *le contexte social* (Smyth *et al.*, 06) et *le contexte personnel* qui traite des sous dimensions du *contexte démographique* (Hupfer *et al.*, 06), du *contexte psychologique* (Kim, 08) et du *contexte cognitif* (Timothy *et al.*, 05), (Tamine *et al.*, 08), (Daoud *et al.*, 09).

Les recherches actuelles en RCI se sont focalisées sur la modélisation et l'exploitation du contexte cognitif, notamment les centres d'intérêt de l'utilisateur. Nous nous proposons de présenter dans cet article une approche de RCI qui intègre trois dimensions à la fois : les centres d'intérêt, la localisation et le temps. Cette orientation est en effet motivée par la prolifération de la technologie mobile (PDAs, téléphones mobiles, ...) qui a rendu l'accès à une grande masse de documents sur le web, possible à toute place et à tout moment. Ce nouveau cadre d'utilisation accentue le besoin et la nécessité de prendre en considération des informations du contexte pour améliorer la précision de recherche. En effet, vu les contraintes et particularités techniques des appareils mobiles (des difficultés de saisie de requêtes, zone d'affichage limitée, ...), nous assistons à une pratique de recherche différente de celle des requêtes traditionnelles de bureau. Des études sur les logs des requêtes des mobinautes (Kamvar *et al.*, 07) montrent que les requêtes des utilisateurs sont plus courtes (donc plus ambiguës), qu'il y a moins de requêtes par session et moins d'utilisateurs qui consultent plus loin que la première page des résultats. De plus, d'après les études dans (Sohn *et al.*, 08), 72 % des besoins informationnels des utilisateurs mobiles sont liés à des facteurs contextuels notamment la localisation et le temps.

RCI dans un environnement mobile

Dans ce cadre, notre objectif est de proposer des techniques permettant de supporter un processus de RCI dans ce nouvel environnement. Plus précisément, nous envisageons d'exploiter d'autres sources d'évidence que le contexte cognitif de l'utilisateur afin de mieux cerner les besoins spécifiques des utilisateurs mobiles. Pour cela, on vise à enrichir le contexte de recherche par des annotations du contexte spatio-temporel, notamment la localisation de l'utilisateur et le temps lors de son activité de recherche. Notre contribution consiste à exploiter à la fois le temps et la localisation de l'utilisateur avec différentes représentations possibles pour répondre à des requêtes sensibles à la localisation et/ou le temps. De plus, nous exploitons le contexte spatio-temporel pour caractériser des situations dans lesquelles se trouve un utilisateur interrogeant le système de RI. L'idée est de construire pour chaque situation identifiée, un profil regroupant les centres d'intérêt appris sur la base des documents consultés dans cette situation.

Cet article est organisé comme suit : la section 2 présente une synthèse de travaux connexes. La section 3 présente notre approche pour la représentation et l'exploitation du contexte dans un environnement mobile. On termine, en section 4, par une conclusion qui résume notre contribution et présente notre perspective de recherche.

2. Recherche contextuelle d'information dans un environnement mobile

Les moteurs de recherche traditionnels considèrent peu le contexte de la recherche et ne sont pas adaptés à l'environnement mobile. Des travaux récents tentent d'améliorer les performances de recherche dans cet environnement. Une première catégorie de travaux a abordé les questions liées à l'adaptation de la recherche aux contraintes imposées par les fonctionnalités limitées des appareils mobiles. Des approches sont proposées pour adapter de la visualisation de la liste des résultats (Sweeney *et al.*, 06) et pour faciliter la saisie des requêtes (Schofield *et al.*, 02).

Une autre catégorie de travaux a porté sur l'exploitation du contexte de l'utilisateur mobile pour améliorer la précision de ses résultats de recherche. Dans (Yau *et al.*, 03) les auteurs appliquent des techniques de data mining sur un historique d'usage, composé de données du contexte, d'actions et de données liées aux actions, pour construire les profils d'un utilisateur qui reflètent ses comportements, ses intérêts et ses intentions dans chaque situation. Quand une requête est émise par l'utilisateur mobile, elle est interceptée et modifiée sur la base de ses profils et du contexte actuel.

Dans (Panayiotou *et al.*, 06) les auteurs exploitent l'importance du temps et de l'expérience (au travail, en vacances, etc) dans la personnalisation d'un portail de recherche de services web pour un utilisateur mobile. Ils proposent de construire un profil dynamique où les centres d'intérêt sont pondérés selon des zones temporelles apprises par l'étude de la routine journalière de l'utilisateur et ses activités dans

Ourdia Boudighaghen

chaque zone. De plus, pour modéliser le changement des préférences de l'utilisateur selon ses expériences, l'association des poids d'importance aux concepts du profil est établie pour chaque nouvelle expérience de l'utilisateur. Ces profils basés temps évoluent et sont maintenus sur la base des feedbacks utilisateurs.

Dans (Hattory *et al.*, 07) les auteurs traitent le problème de l'ambiguïté des requêtes en utilisant une méthode d'expansion basée sur la localisation. Leur méthode consiste à récupérer les coordonnées spatiales de l'utilisateur, les transformer en des mots contextuels (noms de places, d'activités liés à cette place) en utilisant un système d'information géographique et des techniques d'apprentissage des *weblogs*, puis à pondérer ces mots contextuels sur la base de la comparaison de la probabilité globale du mot contextuel dans l'ensemble des documents du corpus cible et sa probabilité locale dans les documents retournés par la requête originale, en fin, la requête originale est étendue par les mots contextuels ayant les poids les plus élevés.

Dans (Ala-Siuru *et al.*, 06) les auteurs présentent une méthode qui combine l'identificateur de la cellule GSM à laquelle est connecté l'utilisateur et les adresses MACs des dispositifs *Bluetooth* à côté, pour caractériser une situation de l'utilisateur mobile. Un raisonnement par cas, peut être conduit alors, sur la base du contexte courant et des situations apprises, pour inférer le profil adéquat du mobile.

Cet article présente notre approche pour une recherche d'information contextuelle dans un environnement mobile. En comparaison aux travaux précédents, notre approche peut être caractérisée par :

- la combinaison de l'adaptation à la localisation, au temps et aux centres d'intérêt de l'utilisateur.
- l'exploitation à la fois de représentations brutes (récupérées des capteurs mobiles) et sémantiques (des concepts d'ontologies) du contexte.
- la construction de profils situationnels à base de l'historique de recherche annoté par le contexte spatio-temporel.

3. Adaptation aux centres d'intérêt de l'utilisateur mobile, à sa localisation et au temps

3.1. Approche générale

Le schéma général de notre approche de RCI mobile est représenté dans la figure 1. L'utilisateur est engagé dans une situation (au travail, fait du tourisme,...) Lorsqu'il a besoin d'information, il soumet une requête à un moteur de recherche traditionnel, qui retourne une liste de documents. Pour adapter les résultats de recherche aux centres d'intérêt de l'utilisateur mobile, à sa localisation et au temps, un processus de contextualisation est mis en oeuvre, il consiste à construire une représentation de ces éléments contextuels de l'utilisateur et des documents. Puis un processus d'appariement permet de définir une fonction d'appariement pour chaque

type de contexte et de combiner les différents scores contextuels. Les documents seront réordonnés selon ce score contextuel en plus de leur score initial. L'utilisateur sélectionne les résultats qu'il juge pertinents pour sa situation, et l'historique est alors mis à jour. Dans la suite, nous explorons les éléments clés de notre approche.

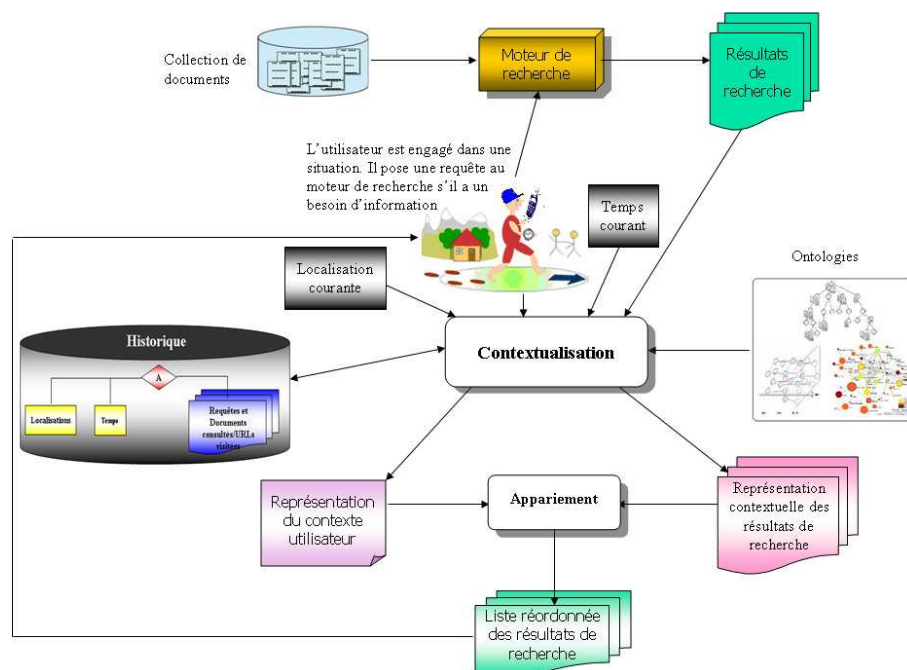


Figure 1. Schéma général de notre approche.

3.2. Processus de contextualisation

Il consiste à créer les représentations du contexte : centres d'intérêt de l'utilisateur, localisation et temps pour l'utilisateur et pour le document.

3.2.1. Les centres d'intérêt

Les centres d'intérêt de l'utilisateur ont été identifiés comme le facteur contextuel le plus important qui réduit la difficulté d'une requête ambiguë dans une tâche de recherche ad hoc (Park, 94). Dans (Tamine *et al.*,08) une distinction est faite entre des centres d'intérêt à long terme désignant les domaines d'intérêt généraux de l'utilisateur, et des centres d'intérêt à court terme reflétant un domaine d'intérêt spécifique relatif à une session de recherche. Notre but est d'affiner la notion de centre d'intérêt vers des besoins situationnels de l'utilisateur. En effet, la mobilité induit des besoins informationnels typiquement conséquents de la situation dans laquelle se trouve l'utilisateur. Nous pensons qu'une partie des informations qui

Ourdia Boudighaghen

aident à clarifier une situation dont découle le besoin en information, sont les aspects contextuels de son moment d'émission, à savoir le temps et la localisation, exemple : "être la nuit à la maison", "l'été à la plage", etc. Cependant, il est clair qu'une situation ne se caractérise pas par des pures coordonnées géographiques et des points temporels. Par exemple la date n'est pas forcément l'attribut à utiliser pour identifier une situation, des attributs tels que le moment de la journée, le jour de la semaine, la saison, se montrent plus pertinents pour la caractérisation d'une situation (le même raisonnement s'applique à la localisation). C'est pour cela que nous proposons d'associer les informations du contexte récupéré par les capteurs du mobile à des concepts sémantiques (extraits d'ontologies temporelles et spatiales) pour pouvoir récupérer toutes les propriétés qui décrivent au mieux les facettes temporelles d'une date et d'une heure et les facettes spatiales des pures coordonnées géographiques. Une situation sera déterminée par une classification sémantique d'une agrégation de points du contexte spatio-temporel. La construction des profils se fera ensuite sur la base de l'historique des recherches attachées aux situations identifiées. Comme dans (Daoud *et al.*, 08), on se basera sur une représentation sémantique des profils par des concepts d'ontologie.

3.2.2. *La localisation*

Pour pouvoir représenter la localisation de l'utilisateur au moment de la requête et caractériser sa situation (nom et type de place), et de plus réaliser une indexation spatiale des documents, un modèle pour la représentation de localisations est nécessaire. L'information géographique peut être représentée à divers niveaux de granularité et sous diverses formes. Pour permettre une bonne représentation de l'information géographique et sa manipulation, la tendance est actuellement vers des approches sémantiques avec des ontologies spatiales. Comme dans (Chen *et al.*, 04) ou dans le projet SPIRIT¹, nous proposons de se baser sur une base de données spatiale et un thesaurus spatial pour représenter et raisonner sur les données géographiques.

3.2.3. *Le temps*

Pour définir les aspects temporels liés au temps de la requête et à la situation de l'utilisateur (matin, soir, *weekend*,...), mais aussi aux documents (heures et/ou dates de disponibilité), un modèle pour la représentation du temps est nécessaire. L'information temporelle est une information complexe, elle est continue et peut être représentée à divers niveaux de granularité. Pour permettre une bonne représentation de l'information temporelle et sa manipulation, la tendance est actuellement vers des approches sémantiques avec des ontologies temporelles. L'ontologie OWL-Time (Pan, 07) est aujourd'hui une référence pour la représentation et le raisonnement sur le temps.

1. Spatially-Aware Information Retrieval on the Internet: <http://www.geo-spirit.org/index.html>

3.3. Processus d'appariement

Une fois que le contexte est représenté, il est exploité dans le processus de RI afin d'améliorer la pertinence de l'information retournée. Notre exploitation consiste à réordonner les résultats de recherche en tenant compte de plusieurs scores contextuels: un score de personnalisation, un score géographique et un score temporel, en plus du score initial d'un document. A chaque fois que l'utilisateur soumet une requête au moteur de recherche, sa localisation et son temps courants sont récupérés. Ils seront utilisés d'une part, pour identifier la situation courante en vue de choisir le profil adéquat pour la personnalisation des résultats de recherche. Et d'autre part, pour calculer un score de pertinence géographique et/ou temporelle des documents retournés.

4. Conclusion et perspectives

Cet article présente une contribution à la définition d'un contexte de recherche dans le cadre d'un utilisateur mobile et son exploitation dans un processus de RI. Un aperçu des approches de la littérature en RCI et mobile est tout d'abord présenté. Ensuite, on a présenté des éléments de notre approche où les éléments du contexte considéré sont explicités avec leur exploitation dans un processus de RI. Nos perspectives de recherche portent en premier lieu sur l'adoption d'une stratégie de caractérisation de situations à partir du contexte spatio-temporel de recherche. En second lieu, sur la proposition d'une méthode de calcul et de combinaison des scores (initial, personnalisation, géographique et temporel) de pertinence des documents.

5. Bibliographie

- Ala-Siuru P., Rantakokko T., «Understanding and recognizing usage situations using context data available in mobile phones», *ubiPCMM06*, 2006.
- Chen H., Perich F., Finin T., Joshi A., «soupa: standard ontology for ubiquitous & pervasive applications», *Int. Conf. on mobile & ubiquitous systems: networking and services*, 2004.
- Daoud M., Tamine L., Boughanem M., «Using a concept-based user context for search personalization», *In Proc of ICDMKE'08*, pages 293-298. IAENG, 2008.
- Daoud M., Tamine L., Boughanem M., Chebaro B. «A Session Based Personalized Search Using An Ontological User Profile». In *ACM Symposium on Applied Computing (SAC 2009)*, ACM, p. 1031-1035, mars 2009.
- Dervin B., Nilan M., «Information needs and uses», *ARIST, William, M.E. Eds*, p. 3-33, 1986.
- Göker A., Myrhaug H., «User context and personalization», in *ECCBR workshop on Case Based Reasoning and Personalization*, Aberdeen, 2002.
- Hattori S., Tezuka T., Tanaka K., «Context-aware query refinement for mobile web search», *Proc. of the 2007 International Symposium on Applications and the Internet Workshops*.

Ourdia Boudighaghen

- Hupfer M., Detlor B., «Gender and Web information seeking: A self-concept orientation model: Research Articles», *Journal of the American Society for Information Science and Technology* 57(8), 1105-1115, 2006.
- Ingwersen P., Jarvelin K., «The Turn: Integration of information seeking and information retrieval in context», *Springer*, 2005.
- Jansen B., Booth D., Spink A., «Determining the User Intent of Web Search Engine Queries», *Proceedings of the 16th international conference on World Wide Web*, 1149 -1150, 2007.
- Kamvar M., Baluja S., «Deciphering trends in mobile search», *Computer*40(8), 58-62, 2007.
- Kim K., «Effects of emotion control and task on web searching behavior», *Information processing and management* 44(1), 373-385, 2008.
- Pan F., «Representing complex temporal phenomena for the semantic web and natural language», Ph.D thesis, University of Southern California, December 2007.
- Panayiotou C., Samaras G., «Mobile User Personalization with Dynamic Profiles: Time and Activity», *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, LNCS 4278, pp.1295-1304.2006.
- Park T., «Toward a theory of user-based relevance: A call for a new paradigm of inquiry». *J. Am. Soc. Inf. Sci.*, 45(3):135-141, 1994.
- Schofield E., Kubin G., «On interfaces for mobile information retrieval», *Mobile HCI 2002*, LNCS 2411, pp. 383-387, 2002.
- Shamber L., «Relevance and information behavior», *ARIST, William, M.E. Eds*, p. 3-48, 1994.
- Smyth B., Balfe E., «Anonymous personalization in collaborative web search», *Information Retrieval* 9(2), 165-190, 2006.
- Sohn T., Li K. A., Griswold W. G., Hollan J. D., «A Diary Study of Mobile Information Needs», *CHI 2008*, April 5-10, 2008, Florence, Italy.
- Sweeney S., Crestani F., «Effective search results summary size and device screen size: Is there a relationship?», *Information processing and management*, 42(4): 1056-1074, 2006.
- Tamine-Lechani L., Boughanem M., Daoud., «Evaluation of contextual information retrieval effectiveness : Overview of Issues and Research », *Knowledge and Information Systems Journal*, to appear 2009.
- Tamine L., Boughanem M., Zemirli N., «Personalized document ranking: Exploiting evidence from multiple user interests for profiling and retrieval», *JDIM*, 6(5), 2008.
- Tao Y., Mamoulis N., Papadias D., «Validity information retrieval for spatio-temporal queries», *Advances in Spatial and Temporal Databases: 8th International Symposium*, (LNCS 2750), 159-178, 2003.
- Timothy M., Sherry C., Robert M., «Hypermedia learning and prior knowledge: domain expertise vs. system expertise», *J. of Computer Assisted Learning* 21(12), 53-64, 2005.
- Xie H., «Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment», *Information processing and management* 44(3), 1346-1373, 2008.
- Yau S., Liu H., Huang D., Yao Y., «Situation-Aware Personalized Information Retrieval for Mobile Internet», *In Proc. of the 27th Annual International Computer Software and Applications Conference*, p. 638-645, Nov, 2003.