Proceedings of the 2nd International Workshop on

Context-Based Information Retrieval

Bich-Liên Doan
Joemon Jose
Massimo Melucci
(Editors)

AUGUST 2007 ROSKILDE UNIVERSITY COMPUTER SCIENCE RESEARCH REPORT #114

Computer Science
Roskilde University
P. O. Box 260
DK–4000 Roskilde
Denmark

Telephone:	+45 4674 3839
Telefax:	+45 4674 3072
Internet:	http://www.ruc.dk/dat/
E-mail:	datalogi@ruc.dk

ISSN 0109–9779

This research report constitutes the proceedings of the *2nd International Workshop on Context-Based Information Retrieval* which is held in conjunction with the 6th International and Interdisciplinary Conference on Modeling and Using Context (CONTEXT 2007), Roskilde University, Denmark, August 2007.

Research reports are available electronically from:

http://www.ruc.dk/dat/

CONTEXT 07

2nd International Workshop on Context-Based Information Retrieval

Held in conjunction with the 6th International and Interdisciplinary Conference on Modeling and Using Context

August 21, 2007 – Roskilde, Denmark

Organized and Edited by

Bich-Liên Doan, SUPELEC, France
Joemon Jose, University of Glasgow, UK
Massimo Melucci, University of Padua, Italy

Preface

With the explosion of the volume of available information on the Web, the information retrieval systems (IRSs) have become necessary to help the users to find relevant documents. Traditional IRSs may be drastically different if they may take into account the context of the information and the context of the query. Contextual information includes both explicit and implicit knowledge about intentions of the user, environment the user is attached to and the system itself. Such factors constrain the search without intervening in it explicitly. One idea to improve the efficiency of IRSs is to make explicit the contextual information in order to enhance the retrieval performance.

The workshop audience is intended to explore related work, theoretical framework and applications which focus on contextual IRSs. It is intended for researchers in the area of computer science, information retrieval, and digital libraries. The workshop's aim is to promote exchanges in these fields, to establish the current state of the art, to identify the emerging problems and to propose future research directions.

This is the second edition of the Context-based Information Retrieval (CIR-07) that is held in conjunction with CONTEXT 07, August 21, 2007 in the University of Roskilde, Denmark.

We thank the program committee members for all the time spent in the reviewing process so that they ensured the high quality of accepted contributions. We thank all the participants for their great interest and effort displayed in preparation of their work. Finally, we thank the organizing committee and local organizers of the CONTEXT-07 conference for their continuous support to preparing the workshop.

Organizers

Bich-Liên Doan, Supélec, France
Joemon Jose, University of Glasgow
Massimo Melucci, University of Padua

Program Committee

Maristella Agosti, University of Padua, Italy
Patrick Brézillon, LIP6, Paris
Bich-Liên Doan, Supélec, France (Co-chair)
Gilles Falquet, University of Geneva, Switzerland
Hideo Joho, University of Glasgow, UK
Joemon Jose, University of Glasgow, UK (Co-chair)
Mounia Lalmas, University of London, UK
Massimo Melucci, University of Padua, Italy (Co-chair)
Iadh Ounis, University of Glasgow, UK
Ian Ruthven, University of Strathclyde, UK
Rohini K.Srihari, University of New-York, USA
Amanda Spink, University of Pittsburgh, USA
Jana Urban, University of Glasgow, UK
Ryen White, Microsoft, USA
Yi Zhang, University of California, USA

Note: These proceedings are a preliminary version of the updated on-line proceedings that will appear in CEUR proceedings (http://ceur-ws.org)

Table of Contents

A Personalized Retrieval Model based on Influence Diagrams

W. Nesrine Zemirli and Lynda Tamine-Lechani and Mohand Boughanem
{nzemirli, lechani,bougha}@irit.fr

IRIT (Institut de Recherche en Informatique de Toulouse) – UMR 5505
118 route de Narbonne
31062 Toulouse cedex 9

Abstract. A key challenge in information retrieval is the use of contextual evidence within ad-hoc retrieval. Our contribution is particularly based on the belief that contextual retrieval is a decision making problem. For this reason, we propose to apply influence diagrams which are an extension of Bayesian networks to such problems, in order to solve the hard problem of user based relevance estimation. The basic underlying idea is to substitute the traditional relevance function which measures the degree of matching document-query, a function indexed by the user. In our approach, the user is profiled using his long-term interests. In order to validate our model, we propose furthermore a novel evaluation protocol suitable for the personalized retrieval task. The test collection is an expansion of the standard TREC test data with user's profiles, obtained using a learning scenario of the user's interests. The experimental results show that our model is promising.

1 Introduction

A key characteristic of most keyword based retrieval models is that the document relevance estimation depends only on the query representation. In recent years, the explosive growth of Web documents makes such basic information searching models less effective [9]. Indeed, different users expressing the same query may have different goals or interests and expect consequently different results. However, most of the basic retrieval models consider that the user is outside of the retrieval process and then provide generic and impersonal results. In order to tackle this problem, personalized information retrieval (IR) is an active area that aims at enhancing an information retrieval process with user's context such as specific preferences and interests in order to deliver accurate results in reponse to a user query. Contexual retrieval is one of the major long term challenges in IR, defined as [1] *combine search technologies and knowledge about query and user context into a single framework in order to provide the most appropriate answer for a user's information needs*.
The goal of this paper is to describe a formal personalized retrieval model able to integrate the user profile in the retrieval process. Our contribution is particularly based on the belief that personalized retrieval is a decision making problem. For

this reason, we propose to apply influence diagrams (ID)[14] wich are an extension of Bayesian networks (BN) [6] to such problems, in order to solve the hard problem of document relevance estimation. ID constitute a theoretical support allowing us to formalize the utility of the decisions related to the relevance of the documents by taking into account the query in one hand, and the user profile in the other hand. A user profile is viewed as a set of long-term interests learned during the previous retrieval sessions [17]. Each user's interest is represented using a term-weighted vector. This representation offers flexibility allowing to plug our model to various learning methods that identify the user's interests. In order to validate our model, we propose first an appropriate framework evaluation based on TREC test collections and then we carry out a series of experiments in order to show it's effectiveness comparatively to a naive Bayesian model.

The remainder of this paper is organized as follows: Section 2 reviews previous work on personalized IR. Section 3 describes our personalized IR model. Firstly, we introduce all the theoritical concepts related to Bayesien networks; secondly, we show the general topology of our ID and then we give the specific details about the quantitative component by means of probability distributions. Section 4 describes our proposed experimental methodology followed by preliminary experimental results that show the effectiveness of our model. Section 5 draws some conclusions and further work.

2 Related work

Traditional retrieval models presupose that the user information need is completely represented by his query. When the same query is submitted by different users, a typical search engine returns the same result regardless of who submitted the query. This may not be suitable for users with different informations needs [2]. To tackle this problem, many recent works use the user's profile features in order to re-rank the documents [16, 7], to refine the query [15] or to adapt the relevance function [4, 11, 5].

In [16], the authors model the user's interests as weighted concept hierarchies extracted from the user's search history. Personalization is carried out by re-ranking the top documents returned to a query using a RSV[1] function that combines both similarity document-query and document-user. In [7] user profiles are used to represent the user's interests. A user profile consists of a set of categories, and for each category, a set of weighted terms. Retrieval effectiveness is improved using voting-based merging algorithms that aim to re-rank the documents according to the most related categories to the query. The profiling component of ARCH [15] manages a user's profile containing several topics of interest of the user. Each of them is structured as a concept hierarchy derived from assumed relevant documents using a clustering algorithm in order to identify related semantic categories. Personnalization is achieved via query reformulation based on information issued from selected and unselected semantic categories. WebPersonae [4] is a browsing and searching assistant based on web usage mining. The

[1] Relevance Status Value

different user's interests are represented as clusters of weighted terms obtained by recording documents of interest to the user. The relevance of a document is leveraged by its degree of closeness to each of these clusters. Recently, extensions of the Page Rank algorithm [11, 5] have been proposed. Their main particularity consists in computing multiple scores, instead of just one, for each page, one for each topic listed in the Open Directory.

The approach we propose in this paper integrates the user's long-term interests into a unified model of query evaluation. Our approach is different from those above in that we attempt to exploit the user's context as an explicit part of the formal retrieval model and not as a source of evidence to re-rank the documents or adapt a basic relevance estimation function. Our goal is to show how user's interests can be explicitely integrated into a unified model in order to evaluate the utility of the decisions related to the statement of relevance of the documents within a query.

3 The model

In our approach, the personalized retrieval process is viewed as a decision making process which estimates the utility of the decisions related to the presentation of documents in reponse to a user's query taking into account the user's interests. The basic underlying idea is to substitute the traditional function of relevance which measures the degree of matching query-document $RSV(Q, D) = p(Q/D)$, a function $RSV(Q, D, U) = p(D/Q, U)$ where $p(A/B)$ is the conditional probability of the event A knowing the event B and U the user model. In our approach, we consider thar the user model is represented using his long-term interests expressed each one with a term weighted vector. Numerous algorithms [13, 12] allow us to build efficiently such models. In order to formalize this relevance function, we propose the use of an extension of BN namely ID. Our interest in ID is motivated by the fact that they constitute a theoretical framework for the decision problem formalization of document relevancy by taking into account the influence of both user's long-term interests and the query submitted. Indeed, there are several properties of ID that make them well suitable for an application in personalized IR. First, it is common practice to interpret the networks' links in a causal manner, a fact that contributes to both a potentially simplified construction process and a more interpretable user model from the user's point of view. Second, ID are able to handle uncertainty in the domain under consideration with regard to arbitrary subset of variables, e.g., users goals, interests, etc.

An ID is a directed acyclic graph that represents a probability distribution. It uses two components to codify qualitative and quantitative knowledge : (a) A directed acyclic graph $G = (V, E)$, where the nodes in $V = \{X_1, X_2, .., X_n\}$ represent the random variables in a domain as documents of collection, terms indexing these documents, the query, and the user's needs and interests; arcs in E encode conditional (in)dependence or influence relationships among the variables (by means of the presence or absence of direct connections between pairs of variables); (b) A set of conditional probability distribution drawn from the graph structure, where for each variable $X_i \in V$ there is a family of conditional propability distributions $P(X_i/pa(X_i))$, where $pa(X_i)$ is any combination of the values of the variables in $Pa(X_i)$ (the parent set of X_i in G). Furthermore, utility values are attached to utility nodes. ID has been explored in structured document retrieval in [3].

The main features of our model are represented in the following section. After presenting the various components of the model, we will illustrate their exploitation during the query evaluation process.

3.1 The diagram topology

Figure 1 shows the qualitative component of our model.

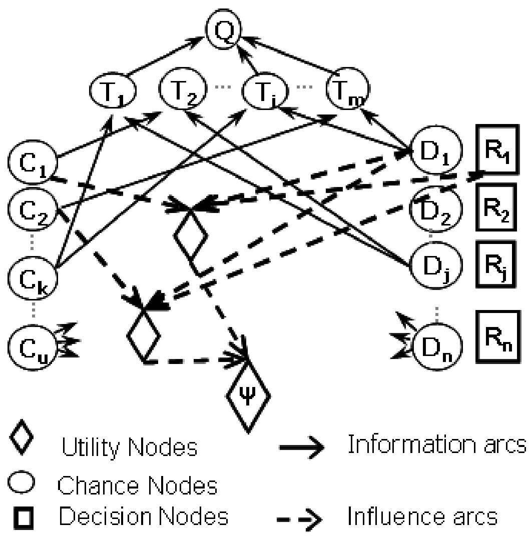


Fig. 1. Influence diagram-based retrieval model

1. **Variables** : The set of variables V is composed of three different types of nodes described below:
   - ***Chance nodes.*** There are four different types of chance nodes $V^{info} = \{Q \cup D \cup T \cup C\}$. The single node Q corresponds to the user's query. It represents the binary random variable taking values in a domain $dom(Q) = \{q, \overline{q}\}$; q indicates that the query Q is satisfied in a given context (related to the user's interests), and $\overline{q}$ indicates that the query is

not satisfied. In our case, we will be interested only by a positive instantiation of Q. $D = \{D_1, D_2, .., D_n\}$ represents the set of documents in the collection. Each document node D_j represents a binary random variable taking values in the domain $dom(D_j) = \left\{d_j, \overline{d_j}\right\}$, where d_j traduces, as in the Turtle model [18], that the document D_j has been observed and so introduces evidence in the diagram, all the remaining documents nodes are set to $\overline{d_j}$ alternatively to compute the posterior relevance. The set $T = \{T_1, T_2, .., T_m\}$ corresponds to the index terms. Each term node T_i represents a binary random variable taking values in the domain $dom(T_i) = \left\{t_i, \overline{t_i}\right\}$, where t_i expresses that *the term T_i is relevant for a given query*, and $\overline{t_i}$ that *the term T_i is not relevant for a given query*. The relevance of a term represents its closeness to the semantic content of a document. The set $C = \{C_1, C_2, .., C_u\}$ represents the set of a specific user's contexts expressing his long-term interests. Similarly, each context node C represents a binary random variable taking values in the domain $dom(C_k) = \{c_k, \overline{c_k}\}$, where c_k and $\overline{c_k}$ express respectively that *the context C_k is observed or not observed for a given query*. The relevance of a user's interest represents its adequacy with the current query.

– ***Decision nodes***. For each document D_j in the collection one decision node R_j is associated which represents the decision to state that the document D_j is relevant with respect to the observed user's interest C_k. The node R_j represents a binary random variable taking values in the domain $dom(R_j) = \{r_j, \overline{r_j}\}$.
– ***Utility nodes***. These nodes express the utility associated to the decision related to presenting the document by taking into account the user's interests. So we associate for each document D_j and each user's interest in the context C_k one utility node. All the values given by the pair (D_j, C_k) are used by a specific utility node in order to compute the global utility attached to the decision to return this document D_j according to the whole user's interests.

2. **Arcs:** The network structure is defined by two kinds of arcs: information arcs and influence arcs.
   – ***Information arcs***. There is a link joining each term node $T_i \in \tau(D_j)$ (terms indexing D_j) to each document node $D_j \in D$ and each context node C_k, whenever T_i belongs to D_j and C_k . This simply reflects the influence between the relevance values of both document and context and term used to index them. There are also arcs which connect each term node with a query node.
   – ***Influence arcs***. These arcs specify the influence degree of the variables associated within a decision. More precisely, in our model, they join the decision nodes, context nodes and document nodes by using an aggregation operator.

3.2 Probability distributions

We will now focus our attention on the probability distributions and the utility values stored in the model. The retrieval inference network is intended to capture all of the significant probabilistic dependencies among the variables represented by the nodes.

- ***Query node***. As previously mentioned, the query is a leaf node that has as many parents as terms are belonging to its representation, noted by $Pa(Q)$. Therefore, it should store 2^k configurations, k being the number of parents. Taking into account only the positive configuration term parents $R(pa(Q))$ (noted further θ), we can compute the probability function attached to a query node using the *fusy-Or* aggregation operator [10] such as:

$$P(Q/pa(Q)) = 1 - \prod_{t_i \in R(pa(Q))} (1 - nidf(T_i)) \tag{1}$$

  where $nidf(T_i)$ is the normalized frequency of the term T_i in the collection.
- ***Term node***. In each term node T_i, a probability function $P(t_i/d_j, c_k)$ is stored. Assuming the independency hypothesis between the document and the user's context, $P(t_i/d_j, c_k)$ is computed as follows:

$$P(t_i/d_j, c_k) = P(t_i/d_j) * P(t_i/c_k) \tag{2}$$

  The probability that a term accurately describes the content of a document and a user's context can be estimated in several ways. We propose the following probability estimation:

$$P(t_i/d_j) = \delta + (1 - \delta) * Wtd(i, j),\ \ \delta \in]0, 1[\tag{3}$$

$$P(t_i/c_k) = \gamma + (1 - \gamma) * Wtc(i, k),\ \ \gamma \in]0, 1[\tag{4}$$

  where $Wtd(i, j) = \frac{wtd(i,j)}{\sum_{t_l \in \tau(D_j)} wtd(l,j)}$ and $Wtc(i, k) = \frac{wtc(i,k)}{\sum_{t_l \in \tau(C_k)} wtc(l,k)}$, $wtd(i, j)$ and $wtc(i, k)$ are respectively the weights of the term T_i in the document D_j and user's interest C_k, δ and γ constant values $(0 \leq \delta, \gamma \leq 1)$.
  More precisely:

$$Wtd(i, j) = 0,5 * \frac{tf_{ij} \log(\frac{N - n_i + 0,5}{n_i + 0,5})}{2 * (0, 25 + \frac{0,75 * dl_j}{avg - dl}) + tf_{ij}} \tag{5}$$

  where n_i is the number of documents indexed by the term T_i, N is the number of documents in the collection, *dl* is the document length and $avg - dl$ the average length of all the documents in the collection, tf_{ij} is the normalized frequency of the term T_i. The context weighting term value $wtc(i, k)$ will be detailed below.

– ***The Utility value.*** As mentioned above, a utility node joins an observed context C_k to the decision related to the presentation of an observed document D_j. According to this, a utility value expresses the degree of closeness between the document D_j to the context C_k. We propose to compute $u(r_j/c_k)$ as follows:

$$u(r_j/c_k) = \frac{1 + \sum_{T_i \in D_j} nidf(T_i)}{1 + \sum_{T_i \in D_j - C_k} nidf(T_i)}, \in [1, 1 + \sum_{T_i \in D_j} nidf(T_i)] \tag{6}$$

We note that the more common specific terms between C_k and D_j there are, the more important $u(r_j/c_k)$ is.

3.3 The query evaluation process

The query evaluation consists in the propagation of new evidence through the diagram, like in BN [6], in order to maximize a re-ranking utility measure. In our approach, this measure is based on the global additive utility value corresponding to the most accurate decisions related to the relevance of a document according to the query and the user's interests. More precisely, given a query Q, the retrieval process starts placing the evidence in the document term nodes then, the inference process is run as in a decision making problem [18], by maximizing the re-ranking utility measure $EU(R_j/Q)$ equivalent to $RSV_u(Q, D)$, computed as follows:

$$EU(R_j/Q) = \sum_{k=1..u} u(r_j/c_k) * P(q/d_j, c_k) * P(c_k) \tag{7}$$

Assuming that pripor probabilities $p(c_k)$ are equal and that documents and contexts are independent, when using the joint law, we obtain:

$$P(q/d_j, c_k) = \sum_{\theta^s \in \theta} [P(q/\theta^s) * \prod_{T_i \in Q \cap (D_j \cup C_k)} P(\theta_i^s/d_j) * P(\theta_i^s/c_k)] \tag{8}$$

Where θ represents the whole possible configurations of the terms in $pa(Q)$, θ^s the s order configuration, and θ_i^s the s order configuration for the term T_i in $pa(Q)$.

Given this latter simplification, the relevance formula (7) becomes:

$$RSV_u(Q/D_j) =$$

$$\sum_{k=1..u} u(r_j/c_k) * \sum_{\theta^s \in \theta} [P(q/\theta^s) * \prod_{T_i \in Q \cap (D_j \cup C_k)} P(\theta_i^s/d_j) * P(\theta_i^s/c_k)] \tag{9}$$

4 Experimental Evaluation

It's well known that the evaluation of an IR model effectiveness is based on using a standard test collection in order to allow accurate comparative evaluation. As example TREC provides widely shared evaluation ressources like test collections and effectiveness metrics in order to evaluate various retrieval tasks like filtering, ad hoc retrieval, web retrieval etc. However, to the best of our knowledge, there is no standard collection for a personalized retrieval task. In order to overcome this limit, we attempt to build a data set wich includes not only testing queries but also user's interests. In the following, we describe how to build such a collection then show the effectiveness of our model.

4.1 Test collection

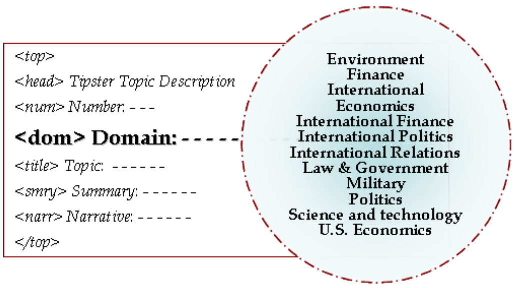


Fig. 2. A TREC query annotated with domain meta data

We used a $TREC$ data set from disc 1, 2 of the ad hoc task, which has a document collection, query topics and relevant judgments. We have particularly used the queries $51 - 100$ as they are enhanced by the domain meta data that gives the query topic. The collection contains queries addressing 12 topics of interest, which are illustrated in figure 2. In order to infer the user's interests, we first applied the following simulation process that builds the trainning data for each domain meta data representing a user's interest:

Once the training (DN_j and DNR_j) set are built for each domain, the related context is built using a long document profile like in the Rocchio algorithm [13], such as:

$$C_k = \frac{\alpha}{|DR|} \sum_{1}^{|DR|} DR_j - \frac{\beta}{|DNR|} \sum_{1}^{DNR} DNR_j, \alpha, \beta \in [0,1], \alpha + \beta = 1. \quad (10)$$

Begin

\# Build a context C_k related to the domain Dom_k

Select a sub set $SubSetQ_k$ **from** $SetQ_k$

For each $q_j \in SubSetQ_k$

$DR_j = \cup_{n=1}^{R_j} \{d_{nj}\}$,

$DNR_j = \cup_{l=1}^{NR_j} \{d_{lj}\}$,

Apply a learning algorithm for each user's interest (DR_j,DNR_j)

End

User's simulation process

Where:

$SetQ_k$: set of queries with the domain meta data Dom_k

DR_j and DNR_j : respectively, the set of relevant and not relevant documents given a query q_j

4.2 The evaluation protocol

In order to evaluate our personalized retrieval model, we compared its performance to a naive bayesian model where the relevance of a document according to the query is computed as follows.

$$P(q/d_j) = \sum_{\theta^s \in \theta} [P(q/\theta^s) * \prod_{T_i \in (Q \cap D_j)} P(\theta_i^s/d_j)] \tag{11}$$

This model represents our baseline. We used the k-fold *cross validation* strategy [8] in our evaluation protocol which simulates the user's interests. For each domain Dom_k of the collection, we randomly divide the query set into k subsets. We repeat experiments k times, each time using a different subset as the test set and the remaining $k-1$ subsets as the training set. This can also be considered as a simulation of user's changing interests as both the training set and the test set change. In addition, the method evaluation is carried out according to the TREC protocol. More precisely, for each query, the 1000 top retrieved documents are first identified. Then, for each value of recall among all the recall points (5, 10, 15, 30,100, 1000), the precision is computed. Finally, the precision is averaged over all the recall points. For the whole data set we obtain a single precision value by averaging the precision values for all the queries. We compare then the results obtained by using our model with those obtained by using the baseline model.

4.3 Preliminary experimental results

The goal of the experiments is to show the effectiveness of our model. All the experiments are carried out with four simulated users, corresponding to the domain meta data presented in Table 1.

Domain meta data	Associeted queries
Environment	59 77 78 83
Law & Government	70 76 85 87
Military	62 71 91 92
Economics	57 72 84

Table 1. The experimented user's interests

In order to estimate the probability distributions associated to the document nodes and the context nodes, we carried out several tunning experiments. The preliminary results allowed us to determine the following parameter values:

$$P(t_i/d_j) = 0,5 + 0,5 * Wtd(i,j) \tag{12}$$

$$P(t_i/c_k) = 0,1 + 0,9 * Wtc(i,k) \tag{13}$$

In the Rocchio learning algorithm:

$$\alpha = 0,75,\ \beta = 0,25 \tag{14}$$

	Naive bayes			Our model		
Queries	P5	P10	Map	P5	P10	Map
57	0,4000	0,6000	0,3311	0,2000	0,4000	0,2457
59	0,2000	0,1000	0,0159	**0,4000**	**0,3000**	**0,0197**
62	0,4000	0,4000	0,2243	**0,6000**	0,4000	0,1833
70	0,6000	0,6000	0,2677	0,4000	0,6000	**0,4147**
71	0,4000	0,2000	0,0569	**0,8000**	**0,7000**	**0,3233**
72	0,0000	0,0000	0,0012	**0,4000**	**0,2000**	**0,0301**
76	0,4000	0,3000	0,0646	**0,8000**	**0,6000**	**0,0878**
77	0,8000	0,7000	0,3990	0,8000	**0,8000**	0,3859
78	1,0000	1,0000	0,7597	1,0000	1,0000	**0,7662**
83	0,0000	0,0000	0,0095	**0,2000**	**0,2000**	**0,0214**
84	0,0000	0,0000	0,0159	0,0000	0,0000	0,0073
85	0,6000	0,8000	0,2170	**1,0000**	0,8000	0,1942
87	0,0000	0,0000	0,0043	0,0000	**0,1000**	0,0041
91	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000
92	0,0000	0,0000	0,0154	**0,2000**	**0,1000**	**0,0221**

Table 2. Experimental results

Table 2, shows the preliminary results obtained using four simulated users. In general we observe that our model gains a statisticaly significant improvement over the baseline at $P5$, $P10$ and mean average precision (MAP). More particularly, our model brings an average improvement of $14,06\%$ in MAP over the baseline accross the whole test queries. However, the increase rate is variable depending generally on the query length. There is also a room for obtaining higher levels of improvement than reported here as we choose reasonable values for a number of parameters (e.g., the weight associated with each term vector representing the user's interests). Future research in this area consists of a much larger scale of experiments as well as an optimization of probability parameters through the exploitation of semantic categories in the context representation, extracted from an ontology.

5 Conclusion

We proposed in this paper a unified model for personalized IR based on influence diagrams, which are Bayesian networks dedicated to decision problems. This model allows to make inferences about the user's search intention and to take ideal actions based on the probability query term distributions over the document collection and the user's contexts represented by his long-term interests. The documents are ranked on the basis of the odd of the utility values correponding to the decisions made on their suitability to the query context.
Furthermore, we attempted to overcome the limit due to the lack of evaluation protocol in our topic area. Indeed, we proposed to augment the widely used TREC test collections by simulated user's interests in order to allow accurate evaluations. The experimental results presented show the effectiveness of our model compared to the naive bayesian one.
In the future, we plan to further the experimental evaluation by experimenting various utility formulations, in particular by identifying other user' contexts parameters to be used for query evaluation.

Ackowledgments

This research was partially supported by the French Ministry of Research and New Technolologies under the ACI program devoted to Data Masses (ACI-MD), project MD-33.

References

1. Allan, J., al: Challenges in information retrieval, langage modeling. Workshop held at the center for intelligent information retrieval (2002)
2. J. Budzik, K.J Hammond, Users interactions with everyday applications as context for just-in-time information access. In Proceedings of the 5th international conference on intelligent user interfaces, pages (44-51), (2000)
3. Luis M. de Campos, Juan M. Fernndez-Luna, Juan F. Huete: Improving the Context-Based Influence Diagram Model for Structured Document Retrieval: Removing Topological Restrictions and Adding New Evaluation Methods. Europeen Colloquim on Information Retrieval (ECIR) pp 215-229 (2005)

4. J.P Mc Gowan , A multiple model approach to personalised information access. Master Thesis in computer science, Faculty of science, University College Dublin, February 2003
5. T. Haveliwala, Topic-sensitive Page Rank,International ACM World Wide Web conference, pp 727-736 (2002)
6. F.V. Jensen, Bayesian networks, decision graphs. Springer. (2001)
7. F. Liu, C. Yu, Personalized Web search for improving retrieval effectiveness, IEEE Transactions on knowledge and data engineering, 16(1), pages 28-40, 2004
8. Mitchell T. M. : Machine Learning. McGraw-Hill Higher Education. (1997)
9. G. Nunberg, As Google goes, so goes the nation. New York Times, May 2003
10. Pearl, J.: Probabilistic reasoning in intelligent systems: networks of plausible Inference. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA. isbn: 0-934613-73-7 (1988)
11. F. Qiu, J. Cho: Automatic identification of user interest for personalized search, International ACM World Wide Web conference, pp 727-736 (2006)
12. Robertson S.E., Sparck Jones K.,: Relevance Weighting for Search Terms, Journal of The American Society for Information Science, 27(3) :(129- 146), (1976).
13. Rocchio, J.: Relevance feedback in Information retrieval. Prentice Hall. Englewood Cliffs. (1971)
14. Shachter, R.D, Probabilistic inference and influence diagrams, Operating Research, (36)4, pp 589-604, 1988
15. A. Sieg, B. Mobasher, R. Burke, Users Information Context: Integrating User Profiles and Concept Hierarchies, Proceedings of the 2004 Meeting of the International Federation of Classification Societies pp 28-40 (2004)
16. Speretta, S., Gauch, S.: Personalizing search based user search histories. In Proceedings of the 13th International Conference on Information Knowledge, Management, CIKM. (2004) 238–239
17. Tamine L., Boughanem M., Zemirli W. N.: Inferring the user's interests using the search history. Workshop on information retrieval, Learning, Knowledge and Adaptatbility (LWA), Hildesheim, Germany, Schaaf, Martin, Althoff, Klaus-Dieter. (2006)108–110
18. H.R. Turtle, W.B. Croft, Inference networks for document retrieval. In Proceedings of the 13_{th} International Conference on Researcch and Development in Information Retrieval (SIGIR), pp 1-24, 1990

Towards an adaptation of semi-structured document querying

Corinne Amel Zayani[1,2], André Péninou[1,2], Marie-Françoise Canut[1,2] , and Florence Sèdes[1,2]

1 IRIT, 118 route de Narbonne, 31062 Toulouse cedex 4, France
2 LGC, 129 A, avenue de Rangueil B.P 67701, 31077 Toulouse cedex 4, France
{zayani, peninou, canut, sedes}@irit.fr

Abstract. *In our research work, we consider that access to semi-structured documents is carried out by a data-oriented query. With different users and a same query, the returned results are always the same although users' characteristics (interests, preferences, etc.) may be different. In order to solve this problem and to offer a personalized access to semi-structured documents, our objective is to improve this type of query in order to adapt the result to each user according to his characteristics. On the one hand, we suggest to reorder the results according to user's interests. On the other hand, we also suggest to establish user's interests implicitly from his queries.*

Keywords: User profile, semi-structured documents, adaptation.

1 Introduction

In order to adapt the content of numeric document, different content adaptation techniques have been defined for different adaptive hypermedia systems such as MetaDoc [1], Plan and User Sensitive Help (PUSH) [2], Hypadapter [3], Personal reader [4]. These techniques are based on rules conceived a priori according to the particular domain.

In our research works, we use semi-structured documents which are stored in centralized documentary repository. These semi-structured documents as well as the returned results by the queries are represented in a logical structure in XML. This structure comprises documentary units (elements of the XML structure). The semi-structured documents can belong to any domain, in this case when different users query this type of documents with a same query, the returned results are always the same although users' characteristics (interests, preferences, etc.) may be different.

On this fact, we aim to automate the content adaptation in generic case. In this paper, we present an algorithm in order to implement another content adaptation technique that is already published. This algorithm allows to reordering the documentary units.

In the suggested approach, in order to reduce the cognitive overload [5] (selection and automatic ordering of the results), the objective is to enrich the query with the user interests determined implicitly [6]. In this paper, we present our research works related to the enrichment of queries in the XQuery language with the user profile in order to adapt the returned results.

The paper is structures as follows. Section 2 presents a general idea of our research works. Section 3 describes the suggested user profile which plays an important role in the process of reordering the documentary units. Section 4 describes the enrichment algorithm of user query. We illustrate in section 5 our research works by an example concerning cottage renting information system in the "Midi-Pyrénées" on the south of France which exists in a platform called PRETI[1].

2 Our proposal

Our context of research works concerns the adaptation of results to the user when he submits queries to semi-structured documents. We work on existing documents which are stored in centralized documentary repository. These semi-structured documents as well as the returned results by the queries are represented in a logical structure in XML. This structure comprises documentary units.

The problem tackled by our work is that documents may not be a priori for adaptation. They are existing semi-structured documents in repository not designed nor built for an adaptation. In such case, different users submitting the same query will get the same results and the same presentation. Because users may have different characteristics, the idea is to present differently the contents for each user when they submit the same query. The objective is to automate this adaptation in order to reduce the cognitive overload of the user when he queries semi-structured documents. We propose an adaptation technique that consists to order the documentary units to be presented to the user according to his own characteristics.

Generally, individual characteristics of users are modeled in user profiles [5], [7]. The characteristics correspond to several information relating to each individual, such as personal information (name, age, etc), interests, preferences, etc. Our objective is to define the interests of each user implicitly (i.e. to determine automatically user interests).

In order to acquire user's interests, we analyze the queries made by the user. In queries, conditions are asserted over elements of semi-structured documents. We consider that conditions may define some user's interest, at a given moment, over the document collection. So, conditions of a query help us to determine user's interests: conditions on elements may become a user's interest. We consider that a condition in a query at a given moment denotes a user's interest. For example, the condition of a query "city=Narbonne" at a given moment denotes the user's interest for that condition. In the user profile, an interest is represented as a condition of a query (an XQuery expression for a condition).

Having such interests, we propose to enrich the user query with some user's interests in order to order the documentary units returned by the query in order to adapt the result to the user.

---

1 www.irit.fr/PRETI

To carry out the exploitation of the user's interests, the enrichment of the query and the update of the profile, we have defined an algorithm [6]. In this paper, we present an improvement of this previously proposed algorithm.

3 User profile

The user profile which we proposed comprises two characteristics (see figure 1.a):

- permanent characteristics which are introduced by the user and which remain fixed in time, such as name, first name, etc.
- changing characteristics which evolve over time. This type of characteristic introduces the user's interests and user's preferences that are determined implicitly without the user intervention. In this paper, we are interested in the user's interests described in XML Schema[2] in the figure1.b.

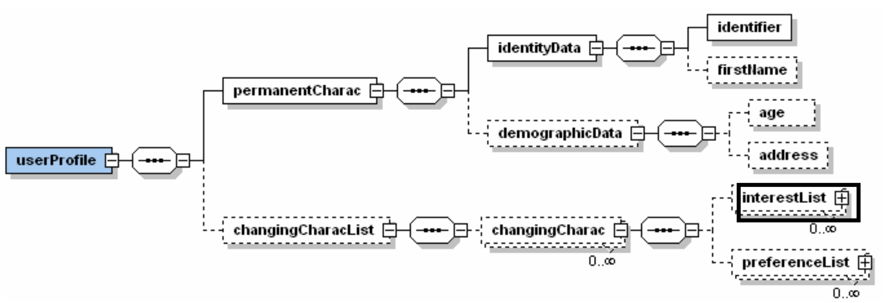


Fig. 1.a. The user profile in XML Schema

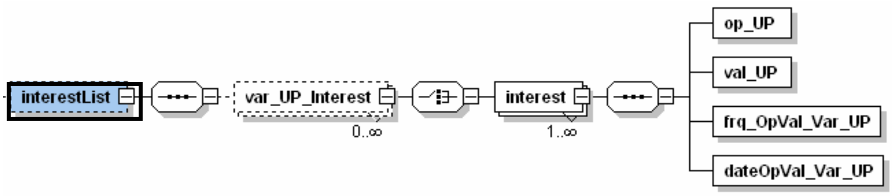


Fig. 1.b. The user's interests in XML Schema

We present below an example of a user profile which represents the instantiation of the XML schema for the PRETI platform (see section 5 for more details). For example, the user is interested in the distance to the sea-distance (var_UP_Interest = sea-distance) which is < (op_UP = "<") to 2 (val_UP = 2) with a frequency of 3 (frq_opVal_Var_UP = 3).

[2] http://www.w3.org/TR/xmlschema-0/

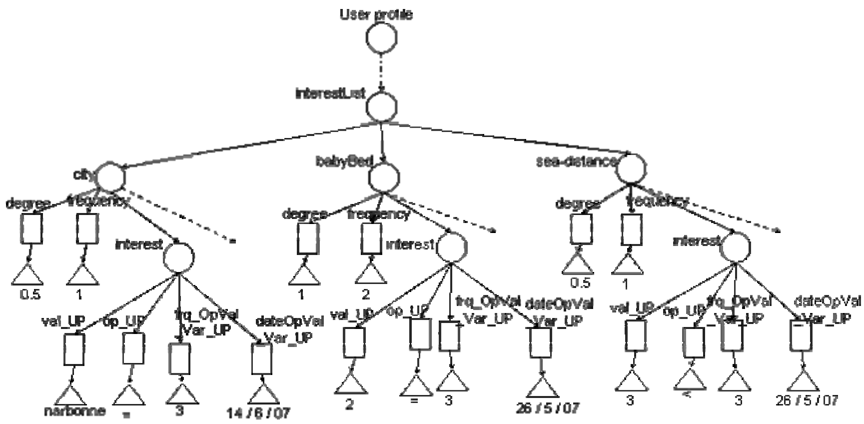


Fig. 2. An example of user's interests

The figure 1 shows the elements that appear in changing characteristics of the user profile. Each element of interest variable "var_UP_Interest" (for example "city", "babyBed" in the figure 2) is associated with an attribute "frequency" that defines "the number of times the interest is used" in the condition part of queries. This attribute "frequency" is incremented each time the element "var_UP_Interest" is used in a query. For example, in the state of the profile of the figure 2, the condition of a new query: "sea-distance<2" will increment the frequency attribute of the "var_UP_Interest" (i.e. sea-distance). According to these frequencies, we proposed to calculate the user's interest degree by the following equation:

$$\text{degree\_interestVal}_i = \frac{\text{frequency\_interestVal}_i}{\sum_{j=1}^{j=n} \text{frequency\_interestVal}_j} \quad (1)$$

In order to standardize the values of interest degree in the interval [0..1], we proposed the following equation:

$$\text{degree\_interestVal}_i\text{\_final} = \frac{\text{degree\_interestVal}_i}{\max_{j=1}^{n} \text{degree\_interestVal}_j} \quad (2)$$

In the equations 1 and 2, "n" represents the total numbers of element "var_UP_Interest" that exist in the user profile.

4 Algorithm for query enrichment

4.1 Principles of the algorithm

For adaptation purpose, the objective of our research work consists in the adaptation of the results during semi-structured documents querying process. For that, we suggest to order (or sort) the documentary units of the results in priority according to user's interests. To carry out this objective, we suggest to enrich the initial query, that is the user's query, with user's interests. Our hypothesis consists in calculating the same returned results in content and length as if no enrichment was made; that is the set returned by the initial query. Only the ordering of the results is modified before presentation to the user. The process should be implemented only after the query has been evaluated by the document repository. However, a query in Xquery language may only return some documentary units of the documents, typically XML elements, meanwhile our process could need to access other elements, typically the entire document. For this reason, we implement an enrichment of user's query which goal is only to order the result, not to change the result in itself. This allows to access any part of the documents when necessary.

We have first proposed an algorithm for query enrichment based on two generic functions in [8]. It takes into account the user's interests correctly if only one interest is given. The two functions suggested have been described in XQuery language [8]. So, we propose to improve this first version in order to take into account "n" interests. As described in section 3, a user's interest is represented as an XQuery condition; for example city="Narbonne", sea-distance<2, …

For taking into account only one user's interest in a given query, the solution consists in ordering results in the following way: 1) those results relevant to the query and to the user's interest, 2) those relevant to the query and *not* to the user's interest. For example, if the user's interest is sea-distance<2, the ordering of the results for a query "Q" (with no conditions on sea-distance) will consist in giving: firstly the cottage having a "sea-distance<2" ("Q and sea-distance<2"), and then the other ones (that is "Q and sea-distance>= 2", or written differently "Q and not(sea-distance<2)"). Without this ordering, the same results would be presented to the user but in some "random" order, regardless of sea-distance value.

When multiple interests are to be taken into account, the ordering problem is to present to the user the same result set in content and length as would give the initial query. Any given result may match no user's interests at all, only ne of them, 2 of them, n of them, all of them. The idea is then to "rank" each result according to its matching to interests in order to get its range in the result presented to the user. That will be a relative range, that is subsets of results matching "n" interests. Finally, results matching all interests will be presented first, then those matching "less" interests, and so on.

For that purpose, we suggest to combine the set of interests and enrich the user query with those combinations. The enriched query will return the results in the correct order. Each interest admits a boolean value when evaluated for a given document. For example, the interest city="Narbonne" will be evaluated to true or false for each cottage renting document (see example in section 5). So, combining interests is a boolean combination of interests and of their negations. For example, combining interests sea-distance<2 and city="Narbonne" leads to 4 expressions: i) sea-distance<2 AND city="Narbonne", ii) sea-distance<2 AND NOT city="Narbonne", iii) NOT sea-distance<2 AND city="Narbonne", iv) NOT sea-distance<2 AND NOT city="Narbonne".

The order of the boolean combination is important since right elements are negated before left ones. So, we use frequencies of interests in the user profile in order to get the right combination that leads to satisfy the "more important" interests first (those having the highest degrees or frequencies), then the lower one, and so on. In the example above, the interest sea-distance<2 is considered as more important than city="Narbonne". We suppose that the user will then prefer cottages that satisfy a sea-distance<2 but not located in "Narbonne" than those located in "Narbonne" but not having a sea-distance<2. Obviously we suppose that he first prefer cottages having a sea-distance<2 and located in "Narbonne", and lastly those corresponding to no interests. The order or stages of the combinations traduces these interests.

Another issue in this approach is to select the user’s interests to take into account for enriching a given query. Using all interests of the user profile regardless of the query may lead to inconsistent conditions and "out of sense" query. So, we try a heuristic solution that is to take into account interests: i) not used in the query condition, ii) having some "sense" in the context of the query. To define a sense relationship between a user’s interest and query conditions, distances measurements inside the document structure between query conditions and user’s interests are used. Heuristic threshold can decide if interests are semantically closed to query conditions, and therefore can be used to adaptation.

Supposing that "n" is the number of the user’s interests that will be extracted from the user profile (and added to the query), the interpretation of the algorithm leads to 2^n stages. Each stage leads to a new "partial" query, that we can consider as a part of the enriched query to be evaluated. At each stage, two parts are inserted in the query being built: a static part and an evolutionary part. The static part is the conditions of the initial query. The evolutionary part is made up of the users’s interests. Its evolution depends on boolean combinations of interests, that is the change of operators (e.g. =, >, <, etc.) by their negation as described above.

In the first stage, we keep the initial operators which are extracted from the profile. That corresponds to find all documents existing in the collection that both match the query and all user’s interests. Afterwards, in each stage, a boolean combination of interests is used. The combination of the interests is made in the decreasing order of frequency existing in the user profile. That corresponds to find documents from the collection that match the query and match the less and less the user’s interests: these documents first match all user’s interests but not the lower frequent, and so on. In the last stage, all operators for user’s interests are the negation in comparison with the first stage. That corresponds to find documents that match the query but that do not match user’s interests.

Finally, in order to adapt the results to the user, the system should evaluate all of these queries, in the order they have been generated. That ensures a correct ordering of the documentary units according to user’s interests.

4.2 Algorithm implementation

The proposed algorithm is divided into three parts:

1) The first part consists in the selection of user's interests to use for query enrichment. This part extracts user's interests from the user profile. A user's interest is selected if: 1) it is not used in the query, 2) it is similar to at least one condition of the query, that is, "it has some common sense and link in the context of the query". Similarity measures are presented hereafter.

2) The second part of the algorithm consists, on the one hand, in sorting the selected user's interests in the decreasing order of frequencies (degrees) and in combining them into a boolean combination as described previously. On the other hand, it generates the enriched query.

To set up these combinations, we use a matrix M [Line, Column] which has a number of columns (Column) equals to the user's interest number (n) and has a number of lines (Line) equals to 2^n. This matrix is filled by values 1 to indicate a user's interest such as stated in the user profile and 0 to indicate the negation of the user's interest. To produce this matrix, we use a function based on classical binary combinations of "n" elements.

We have defined a function called "enrichQuery" (combination_Matrix, initialQuery, interestList). This function takes as parameter the matrix "combination_Matrix", the condition part of the user's query "initialQuery", and user's interests "interestList". This function returns a list named "list_EnrQuery" which contains all the possible combinations of the user interests with the condition part of the user's query.

We have also defined an algorithm using the list defined above ("list_EnrQuery") and able to generate a new query which combines all this sub-queries in a single one to produce the expected results. We use as many let expressions of XQuery language as necessary sub-queries, each one evaluated in the correct order. This generated query is the enriched query evaluated by the documentary repository.

3) The third part of the algorithm makes it possible to update the user profile from the query. This update consists in adding new interests (conditions of the initial query) to the profile or increasing frequencies for existing ones. The execution of the enriched query is carried out via an execution processor SAXON[3]. The JAVA implementation of this algorithm generates only one query in XQuery language submitted to the execution processor.

The first and the third parts of the algorithm are based on the similarity measurements in order to compute the distance between the documentary units and the interests that are in the user profile. The similarity measurements used in our algorithm are inspired by the work of Yi, Huang and Chan [9].

The similarity measurement between the user profile interests and the query conditions is determined on the basis of their elements properties. These elements properties (XML elements properties) are found in the documents themselves. In this paper, we are interested in the similarity measurement for the names of elements taken from [8], that is determined by the name matching of the two elements "a" and "b" as follows:

---

3 http://saxon.sourceforge.net/

$$S_{name}(a,b) = \frac{N_{n(a)\cap n(b)}}{N_{n(a)\cap n(b)} + \alpha(a,b)N_{n(a)/n(b)} + (1-\alpha(a,b))N_{n(b)/n(a)}}$$

Where $N_{n(a)\cap n(b)} = |name(a) \cap name(b)|$ returns 1 if the two names have a common string, else 0. $N_{n(a)/n(b)} = |name(a)/name(b)|$ returns 1 if "a" has a string name difference with "b" string name, else 0. For example, the two names "depart" and "department" have a common string "depart" and the set difference of two names is the string "ment". Therefore $N_{department \cap depart} = 1$, $N_{department/depart} = 1$, $N_{depart/department} = 0$. Supposed that weight α =0.5 (the relative importance of $N_{n(a)/n(b)}$ and $N_{n(b)/n(a)}$ is the same), the similarity of the two names is equal to $S_{name}(department, depart) = 0.66$.
This similarity measurement is used in the first part of the enrichment algorithm, in order to verify if a user interest is used or not in the query conditions. It is also used in the third part of the algorithm to decide for new or existing user's interests.
In our context, we have defined the similarity of parents for "a" and "b" as follows:

$$S_{parent}(a,b) = \begin{cases} 1 \;\; if \;\; S_{name}(parent(a), parent(b)) > 0 \\ 0 \;\; if \;\; S_{name}(parent(a), parent(b)) = 0 \end{cases}$$

This similarity of parents of two elements is used in the first part of the algorithm. It allows to select a user's interest if it has a similarity of parent with at least one condition of the query greater than a threshold. For the moment, this threshold is determined manually and depends on the documents structure.

5 Application

We present, on figure 3, the document structure of the PRETI application[4] in XML schema.

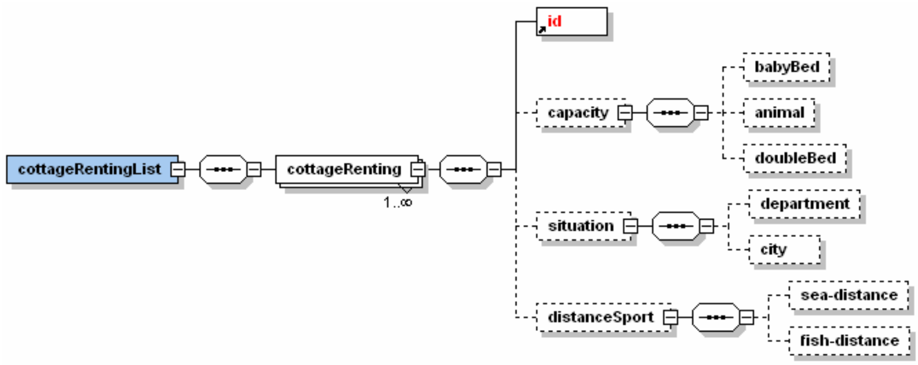


Fig. 3. Example of XML documents of cottage renting

4 www.irit.fr/PRETI

We suppose that a user wants to find rented cottages in "aude" department accepting animals (department = "aude" and animal = "yes"). This query is written in XQuery language as follows:

```
for $a in doc ("cottageRenting.xml")
  where      department="aude"      and      animal="yes"
return $a
```

The result returned by this query contains ordered elements according to the structure of XML document of cottage renting (figure 4).
In figure 4, we suppose that a given user profile is presented on the figure 2 (see section 3) and that the current conditions expressions of query are presented on the figure 4.

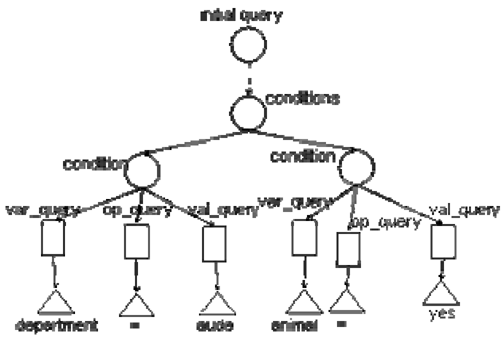


Fig. 4. Example of query conditions

We have applied the algorithm proposed in the section 4. At the beginning, the algorithm determines that:

- the "city" element with value "narbonne" that exits in the user profile has a parent similarity with "department" that exits in the query.

- the "babybed" element that exits in the user profile has a parent similarity with "animal" that exits in the query.

Once these elements are given, the algorithm sorts them according to the decreasing order of their frequencies. In the example, that's give babyBed/city.

On the other hand, the algorithm enriches the query by generating 4 queries (2^2 where the power 2 represents the number of used user's interests). The generated stages in the described example are:

- First stage
  department = "aude" and animal = "yes" and babybed = 1 and city = "narbonne"
- Second stage
  department = "aude" and animal = "yes" and babybed = 1 and city != "narbonne"
- Third stage
  department = "aude" and animal = "yes" and babybed != 1 and city = "narbonne"
- Fourth stage
  department = "aude" and animal = "yes" and babybed != 1 and city != "narbonne"

So the enrichment of the user query with his interests enables to offer to the user a reordering of documentary units of cottage renting.

We have evaluated the results returned by the initial query and by the enriched query with some users of our laboratory. This evaluation is carried out on the corpus of the database PRETI which includes a collection of 700 documents of cottage renting. The result of the evaluation is presented for the first ten documentary units in the following figure:

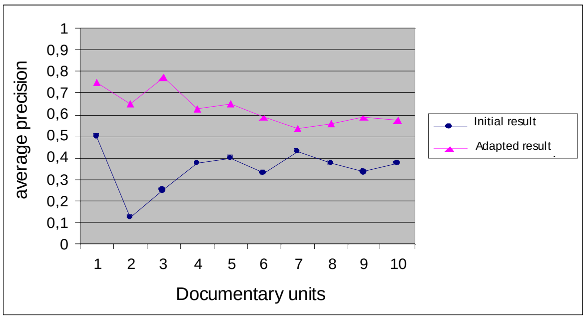


Fig. 5. Evaluation curves with user's judgments

The result of enriched query is presented by the curve of adapted result. The result of initial query is presented by the curve of initial result. The evaluation of these both results in figure 5 shows that the enriched query answers better to user's interests than initial query.

6 Conclusion and prospects

In this paper, we have proposed an improvement of an algorithm for adaptation of results of queries when querying corpus of semi-structured documents. The overall goal is to reduce the problem of cognitive overload. This algorithm consists in selecting the user's interests that are to be taken into account. It also consists in enriching the initial user's query to reorder the results according to the selected interests. A first evaluation with users shows that the first results presented to the user better fits their needs for the enriched query than for the initial users' one. An originality of this work is that this adaptation may be applied to many documents belonging to different domains, especially existing ones where adaptation was not planned when documents were created.

We are now implementing the proposed algorithm in the PRETI platform. This implementation will then enable us to carry out more advanced evaluations. These evaluations must lead to validate our approach from user's point of view. One issue is the number of the stages of our solution in the form of 2^n, n being the number of selected user's interests. One solution could be to modify the enriched query and use the "order by" expressions of XQuery language. Another one could be to give the user a list of subsets of results corresponding less and less to his interests. In that case, only n subsets (n queries) should be evaluated. This possible solutions need to be more deeply studied. Another issue is the selection of user’s interests that are to be taken into account. For that point, we investigate to study other methods from the literature.

Bibliography

1. B. B. Craig D. and A. O. Metadoc : An adaptive hypertext reading system. User Model. User-Adapt. Interact., 4(1) :1–19, 1994.
2. H.-D. B. Hubertus Hohl and R. Gunzenhauser. Hypadapter: An adaptive hypertext system for exploratory learning and programming. User Model. User-Adapt. Interact., 6(2-3), 1996.
3. H. K. Evaluating the utility and usability of an adaptive hypermedia system. In Knowledge Based Systems, 10(5), 1998.
4. P. Dolog, N. Henze, W. Nejdl, and M. Sintek. The personal reader: Personalizing and enriching learning resources using semantic web technologies. In AH, pages 85–94, 2004.
5. P. Brusilovsky. Adaptive hypermedia. User Model. User-Adapt. Interact., 11(1-2), pp. 87–110, 2001.
6. C. Zayani, A. Peninou, M.F. Canut, and F. Sèdes. An adaptation approach: query enrichment by user profile. In Signal-Image Technology and Internet–Based Systems (SITIS), pages 24–35, IEEE/ACM, 2006.
7. Kobsa, A., and Koenemann, J., and Pohl, W. "Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships," in The Knowledge Engineering Review 16(2), 111-155, 2001.
8. C. Zayani, F. Sèdes. An approach to query-based adaptation of semi-structured documents. Dans : International Conference on Enterprise Information Systems (ICEIS 2006), Paphos/Cyprus, INSTICC Press, p. 156-161, mai 2006.
9. S. Yi, B. Huang, and W. T. Chan. Xml application schema matching using similarity measure and relaxation labeling. Inf. Sci., 169(1-2):27–46, 2005.

The role of context in image interpretation

Dag Elgesem[1] and Joan Nordbotten[1]

[1]Department of information science and media studies, University of Bergen
P.O. Box 7800, 5020 Bergen, Norway

{Dag.Elgesem, Joan.Nordbotten}@infomedia.uib.no

Abstract. With the increasing availability and use of mobile phones with camera functionality, it becomes feasible to use these devices to query image databases for information. However, the resulting photo/image query commonly depicts several objects and can be open to a number of interpretations. This leads to ambiguities when determining the intention of an image used as the basis for an information retrieval query. We suggest that this problem is structurally similar to the problem of how to interpret an ambiguous sentence, and that the task can be modeled in a similar way. Though the role of context is a key factor in the solution of the problem of disambiguation of text, we argue that existing accounts of context do not explain what role context plays in image interpretation. In this paper, we propose a definition of image *context* and then show how the disambiguation of images as queries can be modeled as a game of partial information. On the basis of this, a more precise account of the role of context in image interpretation is proposed.

Introduction

Increasing numbers of digitized image collections are available on the Internet and can be accessed via search engines or specialized image retrieval systems. These systems are also available to anyone possessing a mobile phone with Internet connection and camera/image functionality, making image-based query formulation feasible for a broad user community. Unfortunately, current image retrieval algorithms do not yet have the effectiveness of their counter-part text retrieval algorithms, when measured by the degree of relevance of the result sets for the user.

There are 2 main approaches to image retrieval. The most common approach used for Internet access to image databases is a keyword match based on annotations that have been manually defined for each image. This is a tedious task that may not provide good descriptors for the information seeker [5]. An alternative approach, called content-based image retrieval, CBIR, uses an input image as the query statement which is matched to the structural characteristics (color, texture and shape) of the stored images. This approach suffers from the gap between the user's understanding of the semantic meaning of the search image and the current inability of the image retrieval algorithms to identify objects within the image and thus recognize its semantic meaning [1, 5]. Our work[1] is focused on improving the quality of image retrieval based on visual (image) queries by automatically extending the context information in the annotations of image collections and by using context information in the interpretation of visual queries. In this paper we will address the second issue.

An initial problem when an image is used as a query is how context information can be used to determine the user's intention in submitting the image query. In comparison to text queries, an image has no regular structure or components with a defined definition and grammar, and thus far fewer constraints on its interpretation than a string of text.

Consider the following scenario, which could be typical for a tourist with a camera phone. Our tourist is walking around in the city and stops in front of an old church that she finds interesting. There is also a group of sculptures clearly visible above the church door. The tourist wants to know more about the church and pulls out her camera phone, takes a picture of the front of the church, including the sculptures above the church door, and sends it to a multi-modal information retrieval system, MIRS [5] for historical information, expecting to get information about the church. The problem, now, is how can the MIRS determine that she in fact wants general information about the whole church, and not specific information about just the sculptures, when both objects are depicted in the photo?

We will address two general issues in the connection with this question. First, we will suggest that the structure of the problem about the interpretation of the user's intention on sending the picture is analogue to the problem of how an ambiguous sentence is interpreted. We will argue this point in more detail below, but state here that context plays a central role in the process of disambiguation of a

---

[11] Financed by the Norwegian research council, NFR, Project # 176858/S10: Context Aware Image Management http://caim.uib.no/

sentence. This brings us to the second general issue we will address here: what role does context play in the interpretation of an ambiguous sentence, and in the disambiguation of a query in the form of an image?

Let us start with the latter issue. There are a number of definitions of context in the literature (e.g. [2], [3], [6], [8]) and many of them define context in terms of the role the contextual information plays in the interaction between the system and the user. A central example is
Dey's [2] definition of context as

> *any information that can be used to characterize an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves.* (p. 5)

While we agree that Dey's definition is basically correct, it is clear that the reference to what "is considered relevant to the interaction" leaves a lot to be explained. When one considers particular examples of communication it is often easy to point out what information is relevant. But it is important at a theoretical level to explain how and in virtue of what information becomes relevant in a given situation. We suggest that this explanation can be given only by way of an analysis of the structure of communication.

It should be mentioned, however, that Dey does give a partial answer to the question about when information becomes relevant. In his definition of a context aware application he suggests that relevance is relative to the user's task:

> *A system is context-aware when it uses context to provide relevant information and/or services to the user, where relevancy depends on the user's task. (p. 5)*

Again, we share the spirit of Dey's definition but claim that it leaves important questions about how information becomes relevant unanswered. Thus, the definition cannot be made operational. Something more is needed to explain how information becomes contextually relevant.

The problem is that the definition seems to treat the user's task as something that is given. But in the cases we are considering, with ambiguous images, the problem is exactly to figure out what the user's task is. Since context clearly has to be involved in the determination of what the user's task is, we cannot look at the user's task to determine what information is relevant. The system has to use the context to find out what the user's task is. We cannot in general assume that the user's task is known but want to understand how the application can use the context of the communication to determine the right interpretation of the query.

Following the above observations, our suggestion for a definition of context is:

> *Context is the information that must be common knowledge between user and system for communication between them to succeed.*

This definition might not seem radically different from Dey's, but is in fact different in important respects that we will try to make clear through the discussion below.

That this is a reasonable definition of context is one of the points of the paper. The other one, mentioned above, is the suggestion that the problem of determining the user's intention in sending an image to a MIRS is a special case of the problem of disambiguation. We have two arguments for this. The first is that it is intuitively plausible. Consider a person who utters the ambiguous sentence "Every day a man is mugged in Bergen". It seems clear that the most likely interpretation is that "every day there is some man or other who is mugged", rather than "there is one particular man who is mugged every day". How do we know this? Because, first, on the basis of general knowledge about the world. Second, because we can assume that this is common knowledge and that if the speaker had intended the second interpretation he would have used a sentence that was not ambiguous. The point is that the disambiguation cannot happen without bringing in the context. The situation is exactly similar when a statement is made in the form of an image: both the background information about the world and reasoning about what is common knowledge have to be brought to bear in the choice of an interpretation.

Our second argument for the claim that the interpretation of an image is similar to the disambiguation of a sentence is that a model of the latter can be used to analyze the former. In the following we will present and discuss this model of disambiguation.

We are not alone in suggesting that conditions for successful communication are important to the analysis of context and relevance. Mani and Sundaram [6] argue that the key to the understanding of context is to analyze its role in communication. They define context as a “finite and dynamic set of multi-sensory and inter-related conditions that influences the exchange of messages between two entities in communication.” ([6], p. 340) Our approach to the analysis of communication is however different from theirs because we develop our notion of context on the basis of the concept of a game of partial information. The role of common knowledge in the delineation of the context will thus be made explicit in our approach while this is only implicit in [6].

Parikh’s model of disambiguation

There are many pieces of information that might be relevant to determine the user’s context. There is information about location, general background information, information from analysis of the image, etc. But what information is actually useful in the interpretation of a given image-query? Before we can answer this question, an account is needed of the precondition for the common determination of the meaning of an ambiguous query.

In his book, *The Use of Language*, Parikh [7] develops an account of how two communicating agents achieve understanding of the intended meaning of an ambiguous sentence. To this end he uses the framework of games with incomplete information. Applying his theory to our setting, assume that we have a human user U and an automated system S communicating via photographs. The user sends pictures to the system, and the system tries to determine what informational need the image indicates, and sends relevant information back to the user.

The problem can be modeled as a game with partial information. Assume that U moves first and sends a picture pic1, e.g. a picture of a church, to S, and that pic1 has two interpretations ‘church’ (i.e. the whole church) and ‘sculptures’ (i.e. the sculptures on the church wall). The picture is visible to both actors, and hence common knowledge. Assume, further, that U’s intention in sending the picture to S is to communicate the first interpretation, i.e. that she wants to know more about the whole church. But since S does not have direct access to U’s mind, it has to infer it on the basis of general assumptions about the situation and information about the context. There are several further aspects of the context that play a role in the making of this inference, as the model will make clear. Let us explain the details with reference to the following figure:

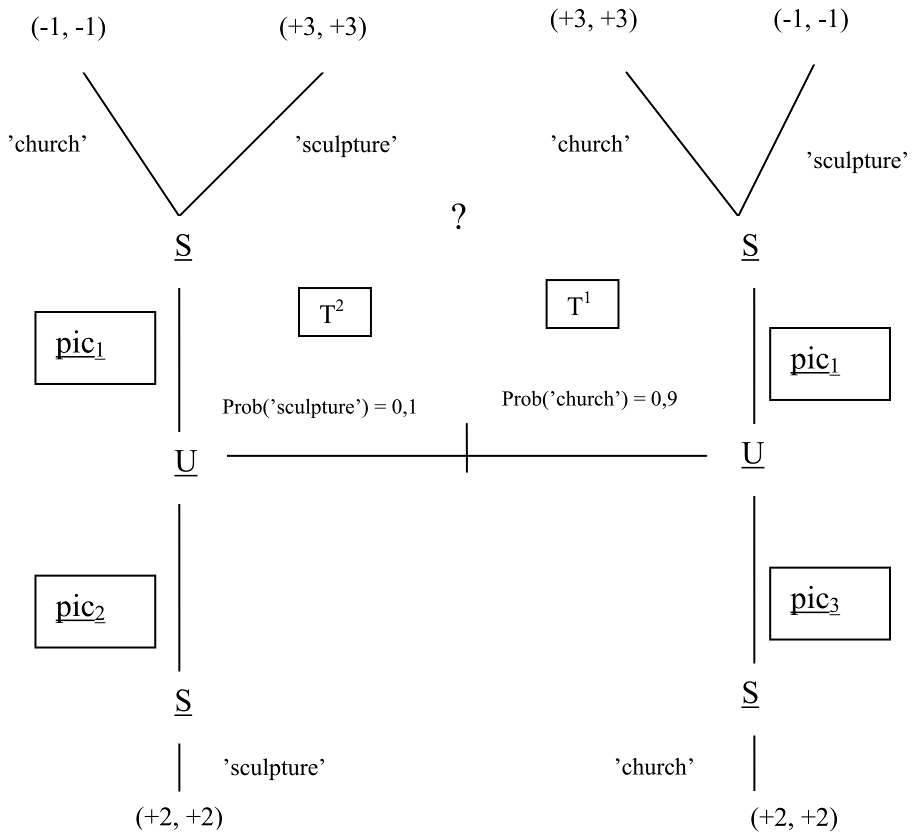


Fig. 1. Image interpretation

By sending the picture, U could either intend to indicate that she is interested in the church or in the sculptures. If the first is the case, U would be in situation T^1. If the intention is to indicate the sculptures, S would be in T^2. The right side of figure is a representation of the situation where U wants to indicate to S that she is interested in the whole church, i.e. T^1, while the left side represents the situation T^2. We see that, as indicated on the central horizontal line, U is more likely (0,9) to want information about the whole church than the sculptures in particular (0,1). This is assumed to be a fact about U at this point of the interaction. (After she has received general information about the church, the probability that she wants information about the sculptures will perhaps increase.) Note that there is a real chance that she would send pic_1 even when she wanted to know more about only the sculptures (i.e. in T^2), hence the ambiguity.

But even though U of course knows her own intentions, S can observe only the picture (pic_1) and cannot be sure whether it is in situation T^1 or T^2. The problem is, again, how can S rationally be sure of U's intention in this game? For this to be possible, two more elements are needed. The first is that knowledge about alternative ways of depicting the object of interest that are not ambiguous. For example, a close up picture of only the sculptures would unambiguously indicate the sculptures. Similarly, a picture taken from a longer distance of the whole church without any surrounding buildings would unambiguously indicate an interest in the church in general. In the figure above, these alternative ways for U to indicate her intention appear in the lower half of the diagram, and are called pic_2 and pic_3, respectively. We see that these alternative ways of depiction have only one interpretation, and one that unambiguously expresses U's intention.

The second element that is needed for S to be able to solve the problem of determining U's intention, is that the parties have to assign values to the possible outcomes, i.e. that a payoff function is defined. There are two factors that will affect the values of the outcomes: the costs of taking the various pictures, and the utility of the chosen interpretation. It is assumed that it is more costly to take a close up picture of a detail than of the whole wall of the building. And, again, it is assumed that a correct interpretation has a positive value, and that misinterpretation has a negative value, for both the user and the system.

To see how this works, consider first the upper, right-hand part of the figure. Here U sends pic_1 to S with 'church' as the intended interpretation. If, now, S chooses 'church' as the interpretation of the picture, this is a positive outcome for both. On the other hand, if S chooses 'sculptures' as the interpretation, we have a case of miscommunication and this would be a negative outcome for both. (Technically, the values assigned to the outcomes could be different, but since the parties are cooperating, it is fair to assume that they have the same valuation.) The situation is different in the upper left part of the figure. Here U sends pic_1 to S with 'sculptures' as the intended interpretation. If S here chooses 'church', this would mean a breakdown of communication and thus a negative outcome, while 'sculptures' would be a positive result for both.

Consider now the lower part of the figure. On the right side, i.e. in situation T1, U sends pic_3 which unambiguously indicates to S that 'church' is the intended interpretation. There is only one outcome and this secures a positive outcome. Similarly on the left side, in situation T^2, where pic_2 unambiguously indicates to S that U's intended interpretation is 'sculptures'. We see that the outcomes in this case, even though they are positive, are valued lower than the positive outcomes in the upper part of the diagram.

This brings us to the second element in the solution of the game of disambiguation of the image. In order to achieve this, U and S "need to compare this ambiguous utterance against an unambiguous one, to ensure that it is more efficient". (Parikh in [7], p. 30) The point is that the outcomes are different because it takes more effort and is thus more costly to create an unambiguous picture. For example, U could have moved up closer to the church and focused only the sculptures (i.e. pic_2). Or, again, she could have moved farther away and taken a picture that captured the whole church and without the sculptures clearly visible (pic_3). This would have communicated that the church was the object of interest. But in both cases the communicative success comes with a price: the extra cost involved in taking more precise pictures. These extra costs are the reason the payoffs are lower in these cases.

The assumptions of common knowledge

With these elements in place, it is possible for a rational agent to determine U's intended interpretation. To reach a unique solution to this problem, a number of requirements have to be met.

1) Both of the agents have to be rational, i.e. their preferences are consistent and transitive, and they maximize outcomes.

2) They have to share a system of ways to depict objects, i.e. there is a language of a sort that is common knowledge.

3) There has to be common knowledge about how structures in the pictorial language refer to objects in the real world.

4) In a given situation there has to be common knowledge of what the possible interpretations are, i.e. that the picture in question (pic_1) is ambiguous. Hence, it is common knowledge that S knows that it is in situation T^1 or situation T^2, but not which.

5) There has to be shared knowledge about how relatively likely the various interpretations are. In our example, it has to be assumed to be common knowledge that the first interpretation ('church') is more likely than the second.

6) The values distributed to the various outcomes by the payoff-function also have to be common knowledge. In our case, this also means that it has to be common knowledge that referring to the objects unambiguously takes greater effort than referring to them ambiguously.

These are of course highly non-trivial assumptions, an issue to which we will return to briefly in the conclusion below. But the important theoretical point for now is that on the basis of these assumptions, it can rigorously be shown that a rational sender will choose the signal that is most efficient and that a rational receiver will end up with the intended interpretation. It is not necessary for our purposes to go into the details of the proof that a unique equilibrium exists, which involves both the idea of a Nash-equilibrium and that of a Pareto-dominance between strategies. The interested reader is referred to Parikh's superb exposition [7].

Context and common knowledge

Parikh's model offers a very powerful account of interpretation and disambiguation of sentences in natural language. The model is helpful with regard to the discussion of the definition of context with which this paper started. The problem with the definition provided by Dey was that it involved essentially the notion of relevance. But this leaves unanswered the question of what information is relevant. But this is exactly what the model discussed above provides. The information that is relevant, and hence makes up the context, is exactly the information that has to be common knowledge in order for communication to succeed.

As the discussion above shows, the framework can be applied to the modeling of the interpretation of images. We believe that even if it abstract, this rational reconstruction of what it is to disambiguate an image is a useful first step in the process of creating a MIRS. It provides a plausible model of what parts of the context that has to be common knowledge in order to secure the communication of the intended interpretation. Hence, the model gives a standard for what the optimal solution to the problem of disambiguation would be and it helps us see where a concrete system is an implementation of a real solution and where we have to make simplifying assumptions.

But we also believe that the framework can in fact provide the basis for the development of a MIRS of the sort indicated in the scenario discussed above. This suggestion gives raise to several questions.

First, images do not depict objects in the same ways as written language does (as established by Goodman in [4]). They do not have syntactic and semantic structures analogous to writing.

A second problem is the amount of world knowledge that is needed to be able to undertake the disambiguation, i.e. to know the set of possible interpretations, their respective probabilities and the valuation of various outcomes.

The third problem is to establish the vast range of common knowledge that is needed in order to solve the problem of disambiguation. This is also a highly non-trivial problem.

A forth problem we will mention is related to the assumption of rationality. It is well known that humans are not perfectly rational in the sense specified by rational choice theory and a context-aware application should be able to take this into account.

Finally all of this information must become common knowledge between the user and the system.

Discussion

We believe that with respect to the class of situations indicated by the scenario above there are practical solutions to all of these problems. First of all, the system we are about to develop is location based in the sense that the images will be tagged with GPS data. It is, of course, possible to automatically recognize objects in images given enough world knowledge in a limited domain. For example, in the CAIM project, mentioned above, we will address this with the use of location data together with image analysis of photographs of a restricted set of historically interesting buildings and objects. Also, an image will contain rich information about the user's point of view with respect to the objects that are the topics of attention. Hence the image that is used as a query will have an indexical function that will limit the set of possible references. In addition, we will be focusing on a very restricted domain where it is possible to provide background information about the kinds of information that are available, i.e. information about objects of historical interest that can be encountered on a walk through the city. Furthermore, the medium itself – the camera phone – put constraint on the range of possible objects. The objects to be communicated about have to be accessible for a photographer. In sum, these constraints will be sufficient to overcome the problems identified above, we believe. This has to be proved in practice, of course, by constructing a working MIRS.

Hence, even if the model cannot be taken directly as a blueprint for the implementation of an application that is able to disambiguate an image, the account still provides important guidance for the development of such a framework. Our claim is that the clearer understanding of the role of context that the game theoretic model provides is a useful basis for the development of context aware image management.

References

[1] Castelli, Vittorio and Lawrence Bergman. 2002. *Image Databases – Search and Retrieval of digital Imagery*. John Wiley.

[2] Dey, Anind K. 2001. Understanding and Using Context. *Personal and Ubiquitous Computing*. **5**:4-7.

[3] Dourish, Paul. 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing*. **8**:19-30

[4] Goodman, Nelson. 1976. *Languages of Art*. Hackett. Indianapolis.

[5] Lu, Goujun. 1999. *Multimedia Database Management Systems*. Artech House.

[6] Mani, Ankur and Hari Sundram. 2007. Modeling user context with applications to media retrieval. *Multimedia Systems*. **12**:339-353.

[7] Parikh, Prasanth. 2001. *The Use of Language*. CSLI Publication. Stanford, CA.

[8] Spink, Amanda and Charles Cole (eds.). 2005. *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer.

A Qualitative Look at Eye-tracking for Implicit Relevance Feedback

Kirsten Kirkegaard Moe, Jeanette M. Jensen and Birger Larsen

Dept. of Information Studies, Royal School of Library and Information Science
Copenhagen, Denmark
kirstenkj@adr.dk, jeanettemjensen@gmail.com, blar@db.dk

Abstract. Our goal in this study was to explore the potentials of extracting features from eye-tracking data that have the potential to improve performance in implicit relevance feedback. We view this type of data as an example of the searcher' immediate context and as containing useful clues of the indications of the interaction between the searcher and the IR system. In particular, we explored if we could qualitatively identify features have potential to improve performance in implicit relevance feedback, and how such features correlate with document elements assessed as relevant or non-relevant. The results point to so-called thorough reading as one of the most promising features for identifying relevant information as input for implicit relevance feedback – in particular when it is related to the total time the searcher has looked an element.

Keywords. Eye-tracking ; Implicit Relevance Feedback ; Thorough reading ; INEX Interactive Track

1 Introduction

The core Information Retrieval (IR) techniques have reached a level of high maturity and do quite a good job of matching document content to a user given query. This is apparent from the widespread use of these techniques in Internet search engines and other search environments. As the core matching techniques have perhaps reached a plateau in terms of performance there has been an increased interest in exploiting the *context* of IR systems more fully to improve search performance[1]. The expectation is

[1] Contextual topics has also appeared at several IR related workshops and conferences recently, e.g., SIGIR 2004 & 2005 workshops on Information Retrieval in Context, the 2006 Information Interaction in Context Symposium, and the Context-Based Information Retrieval workshops at the recent CONTEXT conferences.

that better retrieval performance can be achieved by taking advantage of the context surrounding the IR systems, e.g., from documents or searchers [e.g., 1].

One well-known technique in IR for exploiting the immediate user context is relevance feedback [2], where relevance assessments provided by searchers can be used to modify subsequent queries, e.g., by increasing the weights of query terms found predominantly in relevant documents and decreasing the weights of terms mostly found in irrelevant ones. *Explicit* feedback techniques, based on the active marking of relevant documents, have been studied in some detail, e.g., [3, 4], and generally show that performance gains can be achieved. However, earlier studies have shown that it may be difficult to get explicit relevance assessments from searchers, as the active marking of relevant documents is not a part of the natural workflow in search systems [5, 6]. As a consequence, *implicit* relevance feedback where the feedback data are obtained indirectly from searchers' natural interaction with the system have received increased attention recently. Examples of such contextual behavioural data include: the amount of time that searchers have a document open [7], whether the document is printed [8], or saved [9].

In this paper we work with a type of contextual data that has so far not received much attention for implicit relevance feedback: eye-tracking data of how searchers look at search results. Outside IR research Human Computer Interaction studies have shown that eye-movements can be correlated to human's perception of relevance of read text [e.g., 10, 11]. The studies have, however, been carried out in very controlled (and thus rather unrealistic) environments with quite restricted and simple tasks that are very far from the complexity of realistic information seeking. For instance, in [11] the test persons were asked to identify an answer to a given question from 12 news headlines, or in [10] from 10 sentences. In this paper we investigate if eye-tracking data gathered from a less controlled, interactive IR experiment has potential value as source for implicit relevance feedback. We use a setting where the test persons have a choice of tasks, use a search system similar to an Internet search engine and are free to search and examine any documents in the collection as they wish (see Section 2 for details). In contrast to other studies [e.g., 12] we have chosen to take a qualitative and exploratory approach to identifying potentially useful features from the eye-tracking data rather than an algorithmic one. Apart from not having resources to implement the algorithmic approaches, the main reason is that we wish to study the potentials of eye-tracking data for implicit relevance feedback regardless of whether current algorithmic approaches can identify the observed features. The scope of the study is preliminary and the purpose is to attempt to identify promising features that can be tested empirically in future work. If good performance is obtained with certain features it can then be attempted to implement these algorithmically.

The overall goal of this study is to explore the potential of extracting features from eye-tracking data, regarded here as a type of contextual data, that can improve performance in implicit relevance feedback. Our research questions are:

- Is it possible by qualitative inspection to consistently identify features that have good potential for improving implicit relevance feedback performance from eye-tracking data of an interactive IR experiment?
- How do such features correlate with document elements that have been explicitly assessed as either relevant or non-relevant?

The paper is structured as follows: section 2 gives details of the experimental setting and the methods used in the analysis, section 3 presents the results, and section 4 concludes with a discussion.

2 Experimental setting

The study was carried out as a part of our research group's participation in the INEX2006 Interactive Track experiments [see 13 for more information]. INEX is the INitiative for the Evaluation of XML retrieval which studies the potential of providing more focussed retrieval results (i.e. document *elements*) to searchers by exploiting document structure, e.g., in XML documents [14]. This is mainly done by constructing laboratory test collections. The purpose of the INEX interactive track is to more broadly investigate the behaviour of users when interacting with elements of XML documents, and to investigate and develop approaches for XML retrieval which are effective in user-based environments [13].

In the INEX2006 interactive track the following test material was provided: an element retrieval system backend[2] containing a corpus more than 600,000 XMLified documents from the English version of Wikipedia [15], a prototype element retrieval interface including detailed transaction logging, 12 search tasks, questionnaires and experimental protocols [See 13 for more information]. The test persons acting as searchers were asked to search six of the 12 search tasks (they were given the tasks in pairs and could choose one of them), and were given up to 10 minutes to search for as much relevant information as possible to solve each task. The prototype element retrieval interface (a version of the Daffodil system adapted for element retrieval[3]) displayed the retrieved elements grouped by document, and allowed the searchers to access any full text part of the documents. Searchers could, e.g., access a section directly from the ranked hit list and navigate within the document using an automatically generated table of contents. Searchers were asked to provide explicit relevance assessment of any elements viewed, but were not forced to do so by the system as this might affect their natural interaction behaviour [22]. Assessments could be given using one of five categories [13]:

- **Relevant answer (RA)** – contains highly relevant information, and is just right in size to be understandable
- **Relevant, but too broad (TB)** – contains relevant information, but also a substantial amount of other information

[2] Both an element retrieval backend and a passage retrieval backend were made available. In the present paper we only analyse the tasks searched in the element retrieval backend due to technical problems with the passage backend.

[3] See [16] and http://www.daffodil.de/ for more information on Daffodil.

- **Relevant, but too narrow (TN)** – contains relevant information, but needs more context to be understood
- **Partial answer (PA)** – has enough context to be understandable, but contains only partially relevant information
- **Not relevant (NR)** – does not contain any information that is useful in solving the task

The 'too broad' and 'too narrow' categories are useful when experimenting with elements retrieval systems because they allow searchers to express that a result has some value but an inappropriate granularity.

INEX and its interactive track are particularly interesting in relation to our study: the retrieved and assessed units consisted of parts of documents. This is appropriate in relation to implicit relevance feedback and the eye-tracking data we use as we would typically be able to study patterns of gazing at the level of parts of documents rather than entire documents. In addition, compared to the experimental settings of earlier studies of perception of relevant text in, e.g., [10] or [11] the IENX interactive track is much less controlled and closer to a realistic search situation: the wikipedia corpus is of a general nature that a broad group of searchers would be able to relate to, the tasks were designed to fit the corpus and to be of general interest [20], the test persons could choose between several tasks and were free to interact with the system as they wished – querying, viewing, navigating and assessing any documents or elements from the ranked list, and to stop when they wished. Of course this is still an artificial experimental setting because it was not the searchers' own, real tasks and because the experiment took place in a controlled environment due to the need to use the eye-tracker. In particular, the time limit restriction of only 10 minutes per task is a factor that may affect the results.

In the present paper we analyse data collected from six searchers. In addition to the standard data collected in the interactive track we also collected eye-tracking recordings of all tasks being searched. The Tobii 1750 eye-tracker used provides a large amount of data types, including tracking of the searcher's gaze coordinates recorded at pixel level 50 times per second on both eyes, video screen capture, web cam recording of the test person and logging of keystrokes. As argued above, we have chosen to identify features qualitatively rather than attempting to find useful patterns algorithmically using, e.g., the gaze coordinates as done in some other studies. For our analysis we used a gaze replay visualisation, where the gaze data are overlaid on the video screen capture in real time (See figure 1): A dot shows the current focus on the screen and the trailing line after the dot shows the previous focus. The gaze replay allows us to qualitatively explore any patterns in the searchers' focus on the screen. The interpretation is aided by the web cam recording of the test person ad the tracking status window (Fig. 1).

We limited the analysis to the elements assessed as either Relevant (RA) or Not Relevant (NR). By focussing on the extreme poles of the relevance assessments we hope to get clearer indications that could tell us if some of the observed features can be correlated to relevant and irrelevant elements respectively.

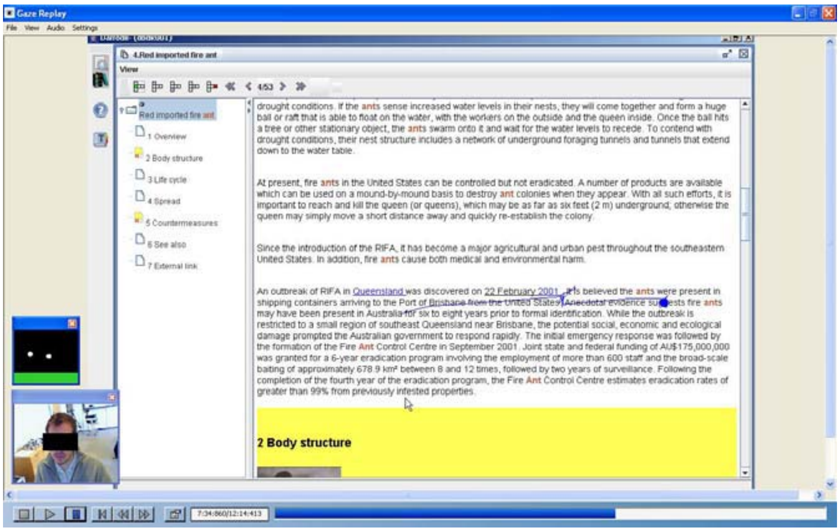


Fig. 1. The gaze replay visualization used for the qualitative identification of features. The blue dot shows the current focus and the trailing line the foci immediately preceding this. A tracker status window with information about eyes and web cam recording of the test person is included to the left.

3 Results

3.1 Overall browsing behaviour

The six searchers completed a total of 18 search tasks in the elements retrieval system. Due to technical problems only 15 of these could be analysed. In these 15 sessions a total of 201 elements were interacted with, and the searchers provided 97 explicit relevance assessments. Searchers were instructed to assess all viewed elements, and this rather low share of assessed elements supports earlier results that it may be difficult to obtain explicit relevance assessments from searchers [5, 6]. The distribution of assessments can be seen in Table 1. In the following only the 33 elements assessed as relevant and the 36 as irrelevant are analysed. In one of the tasks a searcher did not use these categories, which leaves 15 tasks for the analysis.

Table 1. Distribution of the 101 assessed elements.

Assessment	Frequency
Relevant answer (RA)	33
Relevant, but too broad (TB)	17
Relevant, but too narrow (TN)	2
Partial answer (PN)	9
Not relevant (NR)	36

3.2 Identified eye-tracking features

Inspired by existing studies and after an initial screening, where we observed the gaze replay visualisation (Fig. 1) for any gaze behaviour that could be correlated with relevance or non-relevance we choose to focus on three features observed in the gaze replay:

- **Total viewing time**, defined as the total time spent looking at an element relative to the length of the document
- **Thorough reading**, defined as mainly horizontal eye movements, with many fixations per line relative to the number of words on the line and at least half a line read
- **Regressions**, defined as the number of times where searchers regress, i.e., turn back to, an already seen element.

Total viewing time is interesting because some earlier studies have found indications that searchers spend longer time on relevant documents compared to irrelevant documents [e.g., 17, 18]. Kelly & Belkin did, on the other hand, not find any relation between display time and the usefulness of documents [7]. These studies were, however, based on transaction log analysis without the use of eye-tracking, where it was not possible to study if searchers actually did read (or at least looked at) the document content. In our analysis we only include the amount of time actually spent looking at an element. We normalise this with the element length measured in number of lines as it would intuitively take longer time to read a large element than a short and vice versa.

Thorough reading is a central notion because we would expect that any information that has been read, rather than just skimmed or glanced over rapidly, could be useful for implicit relevance feedback. A number of HCI studies have shown that it is possible to differentiate between relevant and irrelevant sentences in text by using eye-tracking [e.g., 11, 19]. Our definition, and our data collection, can be said to be qualitative in that the exact number of fixations in relation to the number of words on the line is not counted – rather it is interpreted qualitatively. We did do some comparisons of inter-coder consistency and found that two coders would agree in the vast majority of cases.

Pfeiffer, Saffari & Juffinger also found that test persons made more *regressions* back to relevant sentences [19]. These studies were carried out in rather restricted settings and with very narrow and simple tasks. In the present paper we use a much more open setting and study if these features can aid also in identifying relevant information in a less controlled information seeking situation with more complex documents and more open tasks.

A number of additional features were considered: skim-reading, skimming and orientation/navigation. It turned out that it was very hard to differentiate between these qualitatively and to define them consistently because they had very short durations. We also attempted to identify thorough reading behaviour from the number of fixations and the duration of these directly from gaze coordinate data, but did not succeed using simple measures.

The results for the three selected features are summarised below.

3.3 Regressions

We counted regressions made back to elements assessed as relevant and irrelevant. As the searchers gaze skip rapidly over the screen when skimming and navigating the documents we counted only those regressions where the searchers had left the element more than one second and the return to look at the element for more than one second afterwards.

A total of 70 regressions were made to the 33 elements assessed as relevant (2.1 per element) in the 15 tasks, and 42 regressions to the 36 irrelevant elements (1.2 per element). That is, we find almost twice as many regressions to relevant elements. However, this is only one more regression per element for relevant compared to irrelevant, and the distribution over searchers is heavily skewed (22 of the elements assessed as irrelevant were given by one searcher). Looking closer at the data no clear pattern emerges from the regressions and based on the present data there is no clear indication that regressions can be exploited to identify relevant elements.

3.4 Total viewing time

We calculated the total viewing time that searchers looked at elements assessed as relevant and irrelevant respectively. Any gazing for more than one second is included, and averaged over the elements. The total viewing time for relevant elements was 43.8 seconds for relevant elements and 9.1 seconds for irrelevant elements, that is, noticeably longer in relevant elements. When we normalise for element length, that is the number of lines in the element, the normalised total viewing time is 3.5 seconds per line for relevant elements on average and 1.1 seconds for irrelevant. Again the distribution is heavily skewed. 5 out of 6 searchers did spend much more time in relevant elements, but one searcher with a large amount of elements assessed as irrelevant spent slightly more time in these. Thus there is a tendency for searchers to spend more time in relevant elements, but it is not unambiguous.

3.5 Thorough reading

In the analysis of thorough reading we have calculated how large a part of the total reading time in seconds that was taken up by thorough reading. On average, the searchers read thoroughly 69 % of the time they spent in relevant elements, and 28 % of the time they spent in irrelevant elements. That is, about 2.5 as much time was spent reading thoroughly in relevant elements compared to irrelevant elements. This varies across test persons, but there is an unambiguous trend that they all spent more time reading thoroughly in relevant elements.

To further analyse this result we have related thorough reading to a contextual task variable given by the INEX interactive track. Each of the 12 search tasks were constructed to be one of three task types (*Decision making*; *Fact finding*; *Information Gathering*) [see 20 for details]. The distribution of thorough reading in relation total viewing time across the three task types can be seen in Figure 2 (note that the percentage can reach 100 for both relevant and irrelevant). In all three cases there is a clear tendency for searchers to spend a larger share of the time reading thoroughly in relevant compared to irrelevant elements. Comparing the tasks types this trend is strongest in *Fact finding* tasks with 77 % thorough reading in relevant elements, and only 15 % in irrelevant. Information gathering also has a large share of thorough reading in relevant elements, 70 % versus 34 % in irrelevant. For these two task types it seems that irrelevant information can be identified fairly easy as elements without much thorough reading, whereas information that searchers need to read more thoroughly tends to be judged relevant. For the Decision making tasks the trend is not so strong with 60 % thorough reading in relevant and 38 % in irrelevant. This is perhaps to be expected as searchers would have had to weight up several alternatives in the process of making decisions.

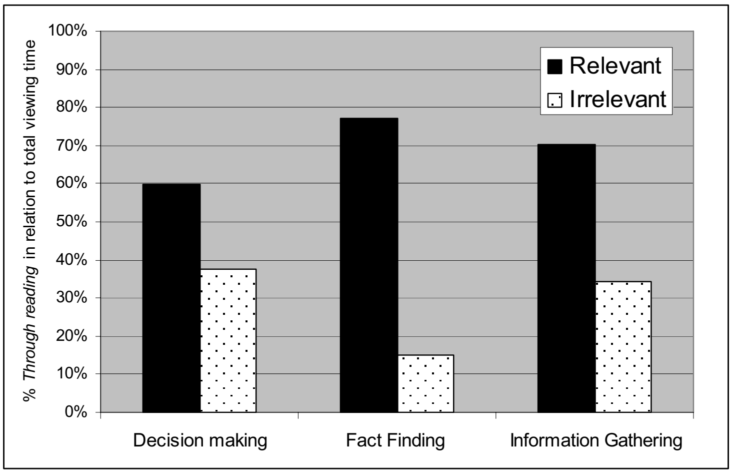


Fig. 2. Share of *Thorough reading* in relation to *Total viewing time* across task types.

4. Discussion

Inspired by features reported in HCI-studies we were able to identify some features by inspecting the eye-tracking gaze replays qualitatively. The features that we could consistently identify were those that were not of very short durations, e.g., less than a second or involving less than half a line of text. Behaviour involving shorter gazing or smaller pieces of text, such as skimming or navigation, proved hard to observe consistently using qualitative inspection.

The feature *thorough reading* was the most promising for identifying relevant information across the six searchers and across task types. Intuitively, through reading corresponds to the notion of having read text as opposed to just skimmed it or glanced over it, and intuitively this would be a good candidate for identifying relevant text.

Our way of operationalising through reading based on qualitative inspection is to the best of our knowledge novel.

However, as some thorough reading did occur also in elements assessed as irrelevant it may be necessary to filter out elements that are read thoroughly, but have a high risk of being irrelevant; including too much irrelevant information in the implicit relevance feedback may seriously harm performance. One way to filter out irrelevant elements could be to set a threshold on the percentage of thorough reading over total viewing time. By ranking the elements by share of thorough reading (not shown) we can observe that the data splits roughly in three bins: the top 33% contains almost only relevant elements, the bottom 33% almost only irrelevant elements and the middle 33% a mixture of both. The implication is that, by setting a threshold of 66% thorough reading out of total viewing time, almost only relevant elements would be included in the relevance feedback. In addition all elements where there was no thorough reading (i.e., corresponding to a threshold of 0 %) were assessed as irrelevant. These elements could thus be used as indications of irrelevance in implicit relevance feedback techniques.

A focus for future research could be to make an algorithmic implementation that can automatically identify thorough reading behaviour. Compared to the other, simpler features analysed in this study (regressions, total viewing time, the number of fixations and the duration of these) thorough reading is a composite concept where several conditions must be satisfied. Thorough reading will thus perhaps take a larger effort to implement, but the gains would also seem to be higher in terms of a better identification of relevant information for the implicit feedback. In addition, the implementation might draw on research already done on reading detection from eye-tracking [e.g., 21], and the output of existing algorithms can be compared to our more qualitative approach. It must be noted that in this study thorough reading has only been analysed in relation to the elements judged relevant or irrelevant. The clear trends shown by thorough reading may be blurred somewhat when thorough reading in elements with intermediate assessments (too broad, too narrow and partially relevant) as well as un-assessed elements are considered.

Exploiting the gaze behaviour of searchers in this manner is a way of bringing the immediate *context* of the user into the IR process: rather than just relying on the user's query to facilitate the match between information need and documents we can attempt to improve the quality of the interaction not only by explicit feedback, but also by implicit feedback from the searcher. The idea of exploiting eye-tracking data can be put in relation Ingwersen's interpretation of the cognitve viewpoint in IR [23-24] and his model of the cognitive communication system for information science [see e.g., 25, p. 33]: by relying on eye-tracking data we are getting indications of the *perception* that takes place as the searcher attempts to understand the information and put it into the context of her knowledge. If any improvement can indeed be obtained by exploiting such eye-tracking data, it may exactly be because they capture indications of the information processing that takes place in the searcher as she strives to make sense of the document. Admittedly, the current cost of an eye-tracker may prohibit this from being applied in any practical settings for some time yet. However, when cheaper eye-trackers (perhaps based on cameras in laptops or cell phones) become available we may begin to exploit this type of context, and the research results produced now, more widely.

5. Conclusions

Our goal in this study was to explore the potentials of extracting features from eye-tracking data that have the potential to improve performance in implicit relevance feedback. We view this type of data as an example of the searcher' immediate context and as containing useful clues of the interaction between the searcher and the IR system that might improve the quality of search results. In particular, we explored if we could qualitatively identify features have potential to improve performance in implicit relevance feedback, and how such features correlate with document elements assessed as relevant or non-relevant. The results indicate that the feature *through reading* have the potential to identify relevant information as input for implicit relevance feedback – especially when it is related to the total time the searcher has looked an element. Theoretically the use of eye-tracking data as contextual clues can be related to the cognitive viewpoint as put forward by Ingwersen [23-24].

Because of the limited size of the study (6 searchers and 15 tasks) the results of the study are indicative only. The size is not only limited because of the available material, but also because of the chosen method: the qualitative identification of features from the eye-tracking data is time consuming and limits the number of search sessions that can be analysed. Nonetheless this explorative approach has allowed us to investigate the value of a number of features without first having to implement algorithms to automatically identify these features. Future research can then focus on those features that show most potential.

We are now working with an extended dataset with 12 searchers, where we have extracted terms from the documents based on total viewing time and thorough reading. The extracted terms will be used in implicit relevance feedback experiments and the performance compared to explicit relevance feedback based on judged elements. The initial results indicate that the implicit relevance feedback generally performs as well as explicit feedback.

Acknowledgments

We wish to thank the test persons who gave their time to participate in the study. We also thank the anonymous reviewers for constructive feedback.

References

[1] Ingwersen, P. and Järvelin, K. (2004): Information retrieval in contexts. In: Ingwersen, P., van Rijsbergen, K. and Belkin, N. eds. *ACM SIGIR 2004 Workshop on "Information Retrieval in Context"*. Sheffield: [University of Sheffield], p. 6-9. [http://ir.dcs.gla.ac.uk/context/, visited 1-5-2006]

[2] Rocchio, J.J. (1971). Relevance Feedback in Information Retrival. In: Salton, G. (red.), *The SMART Retrival System: Experiments in Automatic Document Processing.* Englewood Cliffs, New Jersey: Prentice Hall, 313-323.

[3] Eftimiadis, E.N. (1992): Interactive Query Expansion and Relevance Feedback for Document Retrieval Systems. London, UK: City University. (PhD thesis)

[4] Hawking, D., Thistlewaite, P, and Craswell, P. (1997): ANU/ACSsys TREC-6 experiments. In: Voorhees, E.M. and Harman, D.K. (eds) Proceedings of TREC-6, 275-290.

[5] Beaulieu, M. & Jones, S. (1998). Interactive Searching and Interface Issues in the Okapi Best Match Probabilistic Retrieval System. *Interacting with Computers*, 10, 237-248.

[6] Ruthven, I., Tombros, A. & Jose, J.M. (2001). A Study on the Use of Summaries and Summary-based Query Expansion for a Question-answering Task. In: Thiel, U. & Fuhr, N. (red.), *Proceedings of the 23rd BCS-IRSG European Colloquium on Information Retrieval Research*. Berlin: Springer, 1-14.

[7] Kelly, D. & Belkin, N. (2004). Display Time as Implicit Feedback: Understanding Task Effects. In: Järvelin, K., Allan, J., Bruza, P. & Sanderson, M. (red.), *Proceedings of Sheffield SIGIR 2004: The 27th Annual ACM SIGIR International Conference on Research and Development in Information Retrieval*. Sheffield, UK: ACM Press, 377-384.

[8] Kelly, D. (2005). Implicit Feedback: Using Behavior to Infer Relevance. In: Spink, A. & Cole, C. (red.), *New Directions in Cognitive Information Retrieval*. Dordrecht: Springer, 169-186.

[9] Joachims, T., Granka, L., Pan, B., Hembrooke, H. & Gay, G. (2005). Accurately Interpreting Clickthrough Data as Implicit Feedback. In: Marchionini, G., Moffat, A., Tait, J., Baeza-Yates, R. & Ziviani, N. (red.), *SIGIR 2005: Proceedings of the 28th Annual ACM International Conference on Research and Development in Information Retrieval*. New York: ACM Press, 154-161.

[10] Pfeiffer, M., Saffari A.A., A.R. & Juffinger, A. (2005). Predicting Text Relevance from Sequential Reading Behavior. In: Poulamäki, K. & Kaski, S. (red.). *Proceedings of the NIPS 2005 Workshop on Machine Learning for Implicit Feedback and User Modeling: Workshop at NIPS 2005, in Whistler, BC, Canada, on December 10, 2005.* Helsinki: Helsinki University of Technology, 25-30.

[11] Salojärvi, J., Kojo, I., Simola, J. & Kaski, S. (2003). Can Relevance be Inferred from Eye Movements in Information Retrieval? In: *Proceedings of WSOM `03, Workshop on Self-Organising Maps, September 2003*. Kitakyushu, Japan: Kyushu Institute of Technology, 261-266.

[12] Brooks, P. Phang, K.Y., Bradley, R., Oard, D., White, R. & Guimbretiere, F. (2006). *Measuring the Utility of Gaze Detection for Task Modeling: A Preliminary Study*. Paper presented at Workshop on Intelligent User Interfaces for Intelligence Analysis. Sydney, Australia.

[13] Larsen, B., Malik, S. & Tombros A. (2006). The Interactive Track at INEX 2006. In: Fuhr, N., Lalmas, M., & Trotman, A. (red.), *INEX 2006 Workshop Pre-Proceedings.* Dagstuhl, Germany, 250-261.

[14] Lalmas, M. and Kazai, G. (2006): Report on the ad-hoc track of the INEX 2005 workshop. *ACM Sigir Forum*, 40(1) (To appear)

[15] Denoyer, L. and Gallinari, P. (2006): The Wikipedia XML Corpus. *SIGIR Forum*, 40(1), 64-69.

[16] Malik, S., Klas, C.-P., Fuhr, N., Larsen, B. and Tombros, A. (2006): Designing a user interface for interactive retrieval of structured documents : lessons learned from the INEX interactive track? In: Gonzalo, J., Thanos, C., Verdejo, M. F. and Carrasco, R. C. eds. Research and Advanced Technology for Digital Libraries, 10th European Conference, ECDL 2006, Alicante, Spain, September 17-22, 2006, Proceedings. Berlin: Springer, (Lecture Notes in Computer Science ; 4172)

[17] Morita, M. & Shinoda, Y. (1994). Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval. In: Croft, W.B. & van Rijsbergen, C.J. (red.), *Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: Springer-Verlag, 272-281.

[18] Kim, J., Oard, D.W. & Romanik, K. (2000). *Using Implicit Feedback for User Modeling in Internet and Intranet Searching*. University of Maryland CLIS Technical Report 00-01.

[19] Puolamäki, K., Salojärvi, J., Savia, E., Simola, J. & Kaski, S. (2005). Combining Eye Movements and Collaborative Filtering for Proactive Information Retrieval. In: Marchionini, G. et al. (red.), *SIGIR 2005: Proceedings of the 28th Annual ACM International Conference on Research and Development in Information Retrieval.* New York: ACM Press, 146-153.

[20] Toms, E., Fruend, L. & Jordan, C. (2006). Evaluating Tasks by Type and Form. In: Fuhr, N., Lalmas, M., og Trotman, A. (red.), *INEX 2006 Workshop Pre-Proceedings.* Dagstuhl, Tyskland, 264-266.

[21] Campbell, C.S. & Maglio, P.P. (2001). A Robust Algorithm for Reading Detection. In: *Proceedings of the 2001 workshop on Perceptive user interfaces PUI '01*. New York: ACM Press, 1-7.

[22] Larsen, B., Tombros, A. and Malik, S. (2005): Obtrusiveness and relevance assessment in interactive XML IR experiments. In: Trotman, A., Lalmas, M. and Fuhr, N. eds. *Proceedings of the INEX 2005 Workshop on Element Retrieval Methodology, held at the University of Glasgow, 30 July 2005. Second Edition*. Dunedin (New Zealand): Department of Computer Science, University of Otago, p. 39-42.

[23] Ingwersen, P. (1992): *Information retrieval interaction*. London: Taylor Graham. x, 246 p. ISBN: 0-947568-54-9. [http://www.db.dk/pi/iri/, visited 05.07.2007]

[24] Ingwersen, P. (1996): Cognitive perspectives of information retrieval interaction : elements of a cognitive IR theory. *Journal of Documentation*, 52(1), 3-50.

[25] Ingwersen, P. and Järvelin, K. (2005): *The turn : integration of information seeking and retrieval in context*. Springer. xiv, 448 p. (The Information Retrieval Series ; 18) ISBN: 1-4020-3850-X.

VIVACE Context Based Search Platform

Romaric Redon[1], Andreas Larsson[2], Richard Leblond[1], Barthelemy Longueville[1]
[1] EADS France, Innovation works, TCC5 simulation, IT & systems engineering, Marius Terce. 18,31300 Toulouse, France
{romaric.redon, richard.leblond,barthelemy.longueville}@eads.net
[2] Faste Laboratory, Luleå University of Technology, 971 87 Luleå, Sweden
andreas.c.larsson@ltu.se

Abstract. One of the key challenges of knowledge management is to provide the right knowledge to the right person at the right time. To face this challenge, a context based search platform was developed in the frame of the European Integrated Project VIVACE. This platform is based on the identification of a user context and the subsequent pushing of applicable knowledge to that particular user. We introduce a context model to represent the user's context. This context model is used to describe the context of an engineer working in a specific company. Further, we developed means to index available knowledge based on company engineering context and means to search for knowledge applicable to the user's context. Since it is not always possible to describe in which context the knowledge assets should be applied, we added learning capabilities which enable the system to learn the applicability of specific knowledge to a user's context based on user feedback.

Keywords: Knowledge management, context aware systems and applications, context modelling, analogy and case based reasoning

1 Introduction

This article deals with a presentation of the context based search platform developed in the frame of VIVACE – an Integrated Project within the European Commission's Sixth Framework Programme (FP6). After this introduction Part 2 provides an outline of the VIVACE project and describes the overall objectives of the Knowledge Enabled Engineering (KEE) Work Package. Part 3 focuses on KEE modelling activities: knowledge modelling and context modelling. Part 4 deals with a detailed description of the concept of context based search and a presentation of the platform capabilities. In Part 5 the main features of the prototype developed in 2006 are presented as well as early experimentation results. Foreseen short term and longer term perspectives are discussed in part 6. Finally, part 7 provides some conclusions.

2 VIVACE Knowledge Enabled Engineering

VIVACE [1] is a €70M Integrated Project in the EC Sixth Framework Programme (FP6). The acronym stands for ‘Value Improvement through a Virtual Aeronautical Collaborative Enterprise’, with the main project goal to support the design of a complete aircraft, including engines, by providing increased simulation capabilities throughout the product-engineering life cycle. The Knowledge Enabled Engineering (KEE) Work Package is one of six integrated technical packages. KEE can be considered as the exploitation of Knowledge Management within an engineering context, which fundamentally means leveraging knowledge sources in order to enable engineers to complete their work quickly and correctly. Thus, KEE is about providing the right information to the engineer, at the right time, in the right format, in a collaborative environment that promotes learning within the organization, across the supply chain and across the Extended Enterprise. Therefore, the KEE Work Package proposed to design a context based search platform that would enable users to search for knowledge which is applicable to their contexts.

3 KEE Modelling Activities

In order to proceed towards the development of a platform that provide applicable engineering knowledge depending on the user’s context, a first required step was to specify what we mean by engineering knowledge and user context. The following paragraphs describe the models we introduced in KEE to represent engineering knowledge and user context.

3.1 Engineering Knowledge Modelling activities performed

Engineering knowledge modelling is in itself a wide research area; the purpose of this paragraph is to describe the simple abstract models we used to elaborate our proposal. A more complete description of the different approaches to knowledge modelling could be found in [2].

Engineering knowledge deals with knowledge about products, processes and organisations. A key issue is that engineering knowledge is often stored in people’s head or diluted with other possibly irrelevant, information in technical documents.

In order to support proper capture of engineering knowledge, methodologies such as CommonKADS[3] and MOKA[4] were proposed. These methodologies enable the building of knowledge models composed of interlinked *Knowledge Elements* (K-El). K-El are pieces of knowledge focusing on specific topics. MOKA introduces different forms to support the capturing and structuring of knowledge about both product and process. Entity and Constraints forms enable the collection of knowledge about product breakdown and product limitations. Activity and rules forms enable the collection of knowledge about process breakdown and flow control. At last Illustration forms could be linked to any of the other forms so as to record any corresponding past experience.

These forms are an example of how to organize structured K-El. Other examples of K-El found in industrial companies may include documents about lessons learnt, best practices, expert manuals or also expert contact information etc.

KEE introduces also the concept of *Knowledge Source* (K-Source). A K-Source is a K-El container. Examples of a K-Source could be a simple file repository for managing K-El which are stand alone documents, a web application for managing K-El which are interlinked web pages, or a complex content management system for managing structured interlinked K-El. A K-source usually provides standard capabilities such as index extraction and search capabilities. For a windows file repository, for example these capabilities are provided through Microsoft index server.

3.2 Context Modelling

The objective of this chapter is to describe the work performed in order to represent user context in engineering. The aim was to propose a relevant context model that could easily be understood by engineers and that we can use to quickly develop a platform in order to gain the end-user buy-in.

First of all, it is worth noting that context is still an ill-defined concept such as discussed in [5]. In order to define the context model to use within VIVACE, we investigated two approaches. First a top-down approach which studies existing context models already proposed in the literature and tries to adjust them to fit our needs. Second a bottom-up approach which starts from the study of existing K-El and aims at describing their context of use.

Top-Down Approach. According to Dey et al [6] context is any information that can be used to characterise the situation of an entity. Based on this definition we can propose that engineering context is any information that can be used to characterise the situation of an engineer.

In the literature different context models were proposed, some of them, developed for context representation of mobile users focus more on describing the user's physical context [7] (i.e. his/her current location, device, available resources, etc.), whereas some others include the description of the user's organizational context (role, group membership, tasks, etc.). In this last category, context description model proposed by Kirsch et al. [8] retain our attention. It is based on five viewpoints: space, tool, time, community and process. Several context representation classes are used to describe each viewpoint as shown in the following UML diagram.

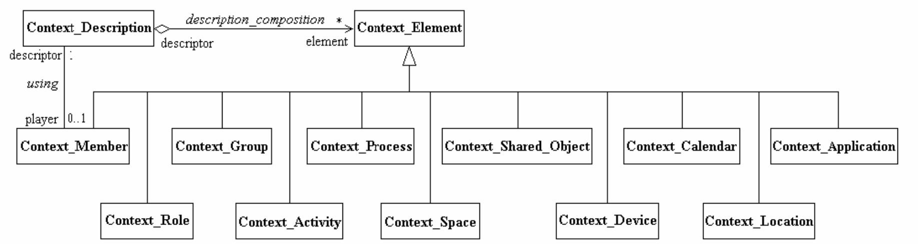


Fig. 1. Kirsch et al. [8] context description model

For KEE use cases, the influence of the user's physical context is not that important: we mainly target engineers working in design offices. Somehow, these engineers always have access to the same resources through the same type of device (i.e. their computer on the network). It means that at a first glance the Context_Space Context_Location, Context_Device representation classes maybe less relevant than the other one for our problem situation.

Bottom-Up Approach. Edmonds [9] says that context is the abstraction of those elements of circumstances in which a model is learnt […], that allows recognition of new circumstances where the model can be usefully applied. In order to achieve our objective of in-context K-El delivery, we focused on representation of an engineering context, which is the abstraction of those elements of circumstances in which a K-El is learnt […], that allows recognition of new circumstances where the K-El could be usefully applied. Therefore, we followed recommendations from Longueville et al. [10] to formalize what they call the explicit context. We studied real examples of K-El coming from VIVACE use cases and we identified how to describe their domain of applicability. The result from this bottom-up approach was the identification of relevant context dimensions. Context dimensions are classes or attributes that describe the context. Six context dimensions arose from the analysis: *product*, *activity*, *project*, *gate*, *role* and *discipline* as shown in the following figure.

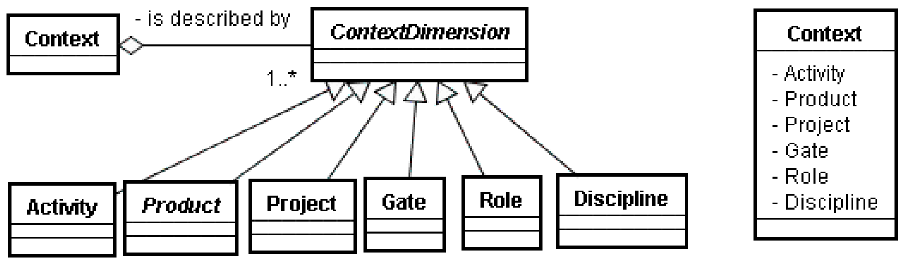


Fig. 2. KEE user context description models: complete and simplified models

Based on this KEE user context description model, and company context dimensions values, the context of user Daniele working in the AVIO company may be described as follows.

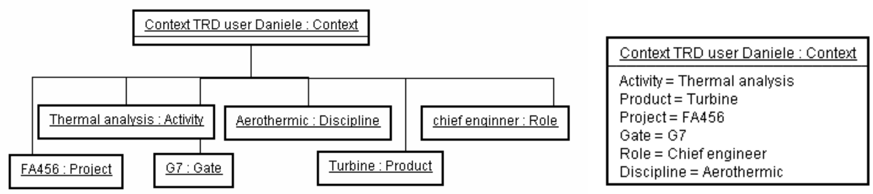


Fig. 3. Description of context of user Daniele working in the AVIO company

4 VIVACE Context Based Search Platform Capabilities

4.1 Overview of VIVACE Context Based Search Platform Capabilities

For the end user, the expected capability of the VIVACE context based search platform is to provide K-El applicable to the user's context described by the following context dimensions: product, activity, project, gate, role and discipline. Therefore, several sub capabilities were developed:

Index K-El on Context. This capability enables the management of the association between a K-El and the different descriptions of contexts in which it was or should be applied.

Context Similarity and K-El Applicability Computation. This capability enables the identification of contexts similar to the user's context, to retrieve K-El which were used in those similar contexts and to compute individual K-El applicability.

At a first glance, these two capabilities may be considered as sufficient for enabling context based search, but since knowledge is not usually indexed on context, the system should include capabilities to learn which knowledge is applicable to which context. Therefore, two new capabilities were developed:

Meta Search in K-Sources. This capability enables searching for K-El in all K-sources through ad-hoc techniques, such as full-text search.

K-EL Applicability Learning and Validation. This capability enables the system to learn that a K-El is applicable to a specific context.

These four capabilities enable a learning process which is shown in the following figure:

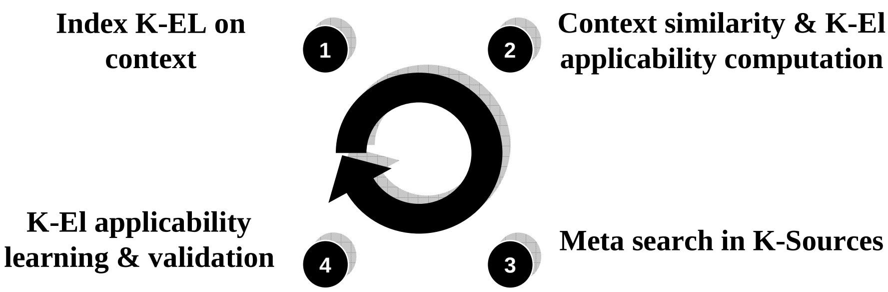


Fig. 4. VIVACE context based search capabilities

The following paragraphs give more detail for each capability

4.2 Index K-El on Context

Indexing K-El on context means to say that a specific K-El is applicable to a specific context with a specific level of applicability. The level of applicability could be an input from the expert or more likely a value computed by the platform itself. In order to index knowledge on context, we specified a *K-El reference and context database* which enables the management of links between contexts and K-El references. A K-El reference is a pointer to a K-El, it contains the K-El identifier and other information such as a title, a description, a type etc.

The platform includes means to specify and deploy K-Source specific extraction and transformation rules that could be launched on a periodic basis. These rules enable interpretation of K-El content in order to generate K-El references and possibly to retrieve context information so as to generate the proper links in the *K-El reference and context database*.

The conceptual data model of the *K-El reference and context database* is shown in the following figure. This model enables i)to manage many to many relationships with associated applicability between K-El references and contexts, ii)to make the distinction between applicability given by an expert and applicability computed/learnt by the system.

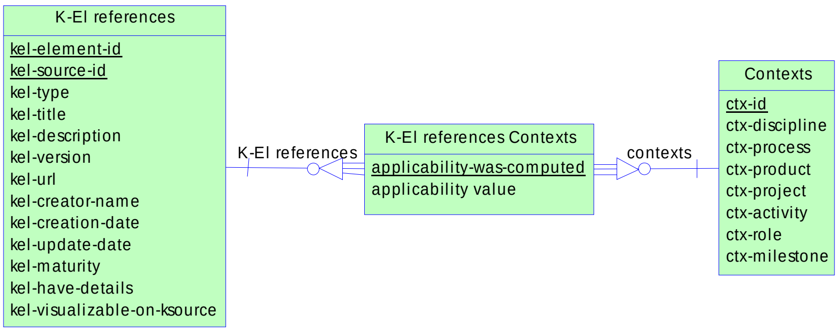


Fig. 5. *K-El reference and context database* conceptual data model

4.3 Compute Context Similarity and K-El Applicability

This is the core capability of the platform; it is based on Case Based Reasoning (CBR) technology [11]. In order to develop our CBR application we had to define the case model, the case base, the viewpoints including associated similarity measures and retrieval strategy, and the adaptation strategy.

Case Model. For our CBR application, a case is composed of i) the description of a user context modelled according to KEE user context description model (problem descriptors) ii) the list of K-El associated to this context with their applicability (solution descriptors). Problem descriptors are considered as a symbol, an ordered symbol or a taxonomy, as described in the following figure:

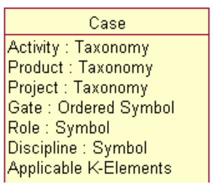


Fig. 6. Case model

Case Base. The case base is obtained from the *K-El reference and context database* and it contains the description of all the contexts with associated K-EL.
Viewpoints. Viewpoints contain information about weights and similarity measures to use for each descriptor as well as for selection of the retrieval strategy. For our application, default weights were proposed; similarity measures such as taxonomy and similarity matrices were used and nearest neighbour retrieval method was selected.
For example the following similarity matrix is used for computing the local similarity for the gate descriptor.

Gate	A1	A2	B4	B5	C6	D7	E8	F9	G10
A1	1								
A2	0,92	1							
B4	0,72	0,92	1						
B5	0,23	0,42	0,62	1					
C6	0,09	0,23	0,42	0,92	1				
D7	0,03	0,09	0,23	0,72	0,92	1			
E8	0,01	0,03	0,09	0,53	0,72	0,92	1		
F9	0	0,01	0,03	0,39	0,53	0,72	0,92	1	
G10	0	0	0,01	0,03	0,09	0,23	0,42	0,62	1

Fig. 9. Similarity matrix used for the gate descriptor

Adaptation. The result of the nearest neighbour retrieval is a list of similar contexts with associated K-El and their applicability. Depending on i) the frequency of occurrences of a specific K-El in this list of similar contexts, and ii) the K-El applicability for each similar context; the platform computes individual K-El applicability to user's context.
The following figure is a simplified representation of the context based search mechanism.

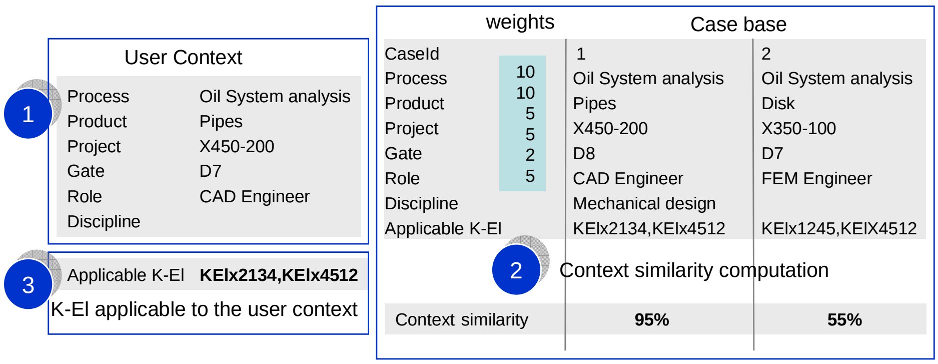


Fig. 10. Simplified representation of the context based search mechanism

4.4 Meta Search in K-Sources

This capability propagates a full text search request to all K-sources connected to the platform. Then, each K-source relies on its own search capabilities to perform the search (for example Documentum™ based K-sources which will use Documentum™ search capabilities and for WWW K-source search will be performed by ad-hoc web search engine such as Google™. The Results of each K-source search process are then sent back to the platform which aggregates results and possibly adds new K-El reference to the K-El reference and context database.

4.5 Knowledge Applicability Learning and Validation

Users can assess K-El applicability to his/her context. For K-El that result from full-text search, applicability is unknown and the user could decide to quantify applicability to his/her context by valuing a percentage from 0% (not applicable) to 100% (fully applicable). For K-El that result from contextual search, the user could decide to increase or decrease the applicability value which was computed by the system.

The platform should include a validation process to control the user's feedback process. This validation process may be implemented differently from company to company. For example, some companies may wait for 5 users to give feedback in the same direction to automatically validate, whereas others may ask an administrator to validate manually on a periodic basis.

After validation, the *K-El reference and context database* is updated according to users' feedback thus enriching potential results of any later context based search.

5 VIVACE Context Based Search Platform Prototype

5.1 Overview of Platform Architecture and Implementation

A first prototype of the platform was developed in 2006 based on a three-layered architecture:

Portal Layer for enabling a user's context identification and contextual and full text searches. The portal implementation is based on portlet technologies, thus it could be easily integrated into other web applications, such as existing company intranets.

Kernel Layer for enabling context similarity and K-El applicability computation as well as managing the *K-El reference and context database*. The kernel implementation is based on EADS Innovation Works CBR engine.

Knowledge Source Interface (KSI) Layer for enabling extraction and alignment of K-El metadata and multi-source search capabilities. The KSI implementation is based on Documentum ECIS services.

The three different layers were based on open source components (APACHE components and MySQL) and they communicate through web services.

5.2 Platform Prototype User Interface

The platform prototype user interface is composed of five main panels:
Context Identification Panel. This panel enables the identification of the user's context through setting appropriate context dimensions values.
Search Panel. This panel enables the launch of the contextual search, which may be combined with complex filter-based K-El metadata and meta search in K-sources.
K-El Browser. This panel enables the browsing of search results

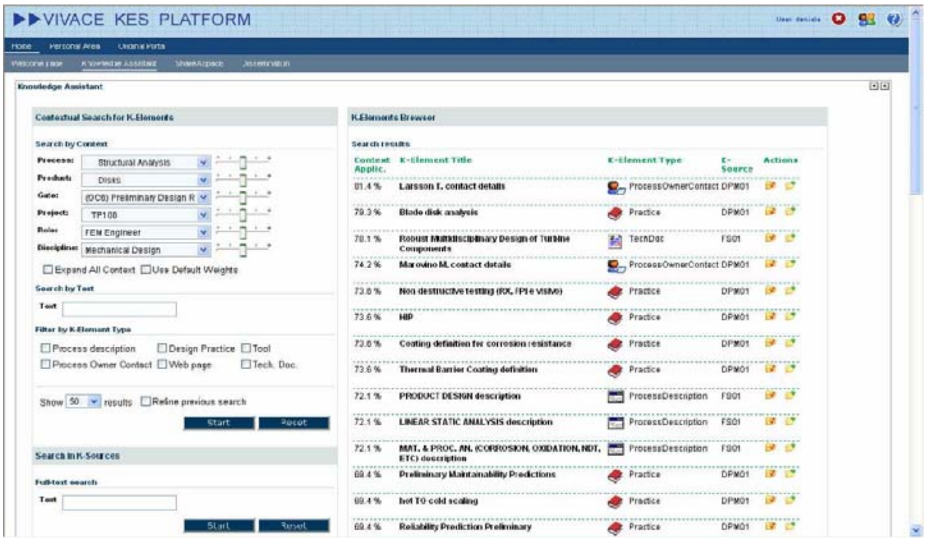


Fig. 7. Search and K-El browser panel

K-El reference and Context Viewer. This panel enables the visualisation of K-El reference and the context in which the K-El are applicable.
K-El Applicability Feedback Form: This panel, which is activated each time the user opens a K-El, enables the user to assess the applicability of the K-El with regard to his/her context.

5.3 Platform Prototype Scenario of Use

This generic scenario provides a walk through of how we envisage the platform being used.

1. User A is working in context Cx
2. Platform provides applicable K-El to user A through context based search.
3. User A accesses K-El reference and he/she could decide to open the K-El. For each opened K-El, the platform requests his/her feedback on K-El applicability.
4. User A is not fully satisfied by K-El obtained in step2, he/she searches for other K-El through a full text search in all K-sources.

5. User A find interesting K-El and he/she records applicability of this K-El to his/her context.
6. User B is working in a context Cy similar to Cx
7. Platform provides applicable K-El to user B. New K-El which were found in step 4 and said to be applicable by user A in step 5 are automatically provided to user B.

5.4 Platform Prototype Experimentation

Based on the VIVACE Turbine Rotor Design (TRD) use case, successful experimentations were conducted by KEE partners to validate the prototype. AVIO experts were involved in the project, they participated in the specification of the TRD context description model and they provided knowledge about similarities between context dimensions values. AVIO internal K-sources and external K-sources, such as the World Wide Web, were connected to the platform.

Results of these first experimentations validate that the platform can help provide the user with applicable knowledge depending on his/her context. Furthermore, promising results were obtained for the indexing of web pages based on context. The platform promotes collective learning about which information available on the web is applicable to which context. Thus, the platform offers promising capabilities to solve the information overload issue that users encounter in engineering activities.

6 Perspectives

6.1 Future Work

Piloting Activities. A piloting phase is scheduled in early 2007 in order to validate the platform in a real industrial environment and to measure associated benefits. The platform will be used by an operational team working on Turbine Rotor Design at AVIO.

SWOT analysis. SWOT analysis aims at evaluating the Strengths, Weaknesses, Opportunities and Threats involved in our proposal. This analysis will rely on consolidated results obtained from the piloting activities and analysis of other proposals for enabling context based search such as for example, those proposed by Kirsh-Pinheiro et al.[12] and David Leake et al.[13].

Context Modelling Enhancement. As described in part III, we based our work on a simplified context description model. The objective is to enhance this context description model in order to take into account i) richer information about existing context dimension (eg context dimension classes rather than attributes, for example, the discipline context dimension class may be described by several attributes such as the name of discipline, the level of expertise, etc.) ii)new context dimension, for example to better describe user profile and tool used iii)latest research results on context modelling.

Context Similarity Computation. The objective is to refine the context similarity computation algorithm to cope with the enhanced context description model and lack of knowledge about local similarities.
Guidelines Elaboration. As software platform or tool alone will never be an answer to a knowledge management issue, appropriate guidelines focusing on organizational, methodological and behavioural aspects should be elaborated. These guidelines will be used together with the platform to address the challenge of in-context knowledge delivery.

6.2 Research perspectives

Advanced Context Identification. The objective is to enable automatic or assisted identification of user context. A user's context may be identified through monitoring and analysis of user behaviours, application and data used etc. However, it may not be possible to identify all context dimensions automatically, so the user may still have to set some context dimension values and validate the identified ones.
Knowledge Pushing. The objective is to combine advanced context identification capabilities and context based search capabilities to automatically push applicable knowledge to the user depending on his/her context. In other words, the aim is to develop a pro-active search system that does not necessarily require the user to take the search initiative.
Learning Enterprise. The objectives are twofold: at the software level to enhance and develop users' feedback mechanisms in our platform; at the organizational level to promote a learning culture in which users are eager to provide their feedback for the benefit of others. Nowadays, with participative tools associated to the Web2.0 framework, an efficient learning organization emerges on the web. Enabling this transformation in industry is still a challenge.

7 Conclusion

In order to face the new competitive situation in industrial companies, the design cycle must be shortened and engineers are asked to design right first time. On the one hand, shortening the design cycle leaves less time for the engineer to search for knowledge, on the other hand, the requirement to design right first time increases the need for knowledge search and reuse. The key issue is then to provide the right knowledge to the right user at the right time in the design process.

In order to face this issue we proposed a context based search platform that enables in-context delivery of knowledge. First results of platform experimentation are very promising. The platform enables collective learning of which knowledge is applicable to which context and efficient searching for knowledge applicable to the user context.

We believe that this context based search platform is a first step toward the development of pro-active search systems. Therefore further research work is required in order to develop automatic context identification capabilities and to use them together with context based search capabilities. Nowadays, engineers are often not even aware that there may be some knowledge available to help them, so they do not take the search initiative. For this reason, pro-active search systems that may push applicable knowledge without requiring user initiative are seen as the ultimate answer for supporting engineering activities.

Acknowledgements. The VIVACE integrated project is partly sponsored by the Sixth Framework Programme of the European Community (2002-2006) under priority 4 “Aeronautics and Space” as integrated project AIP3 CT-2003-502917. VIVACE context based search platform was developed by AVIO, EADS IW and Xerox.

References

1. Value Improvement through a Virtual Aeronautical Collaborative Entreprise.: www.vivaceproject.com
2. Motta, E. Reusable components for knowledge modelling. IOS press (1999).
3. Schreiber, G.,Akkermans, H., Anjewierden, A., de Hoog, R.,Shadbolt N.,Van de Velde, W., Wielinga, B.: Knowledge Engineering and Management: the CommonKADS Methodology. MIT Press, (2000).
4. Callot, M.,Oldham, K.,Stokes, M.,Godwin, N.,Brimble, R.,Klein, R.,Sellini, F.,Merceron, F.,Danino, D.: MOKA - Methodology and Tools Oriented to Knowledge based Engineering. Project EP25418 Public report N°2. (1999)
5. Bazire, M., Brézillon, P.: Understanding Context Before Using It. CONTEXT 05. LNCS 3554. Springer Verlag (2005)
6. Dey, A.K, Abowd G.D.:Toward a Better Understanding of Context and Context Awareness.(1999)
7. Lelouma T.,layaïda N.: Context Aware Adaptation for Mobile Devices, (2004).
8. Kirsch-Pinheiro, M.,Villanova–Olivier, M., Gensel, J., Martin, H.: Representing Context for an Adaptive Awareness Mechanism. CRIWG 2004. LNCS 3198. Springer-Verlag.(2004)
9. Edmonds, B.: The Pragmatic Roots of Context. (1999).
10. Longueville, B., Gardoni, M.:A Survey of Context Modeling : Approaches, Theories and Use for Engineering Design Researches. CONTEXT 03. LNCS 2680. Springer Verlag(2003).
11. Watson, I.,Marir, F. Case based Reasoning: a review. The knowledge Engineering review. (1994)
12. Kirsch-Pinheiro, M.,Villanova–Olivier, M., Gensel, J., Martin, H.: Context-aware Filtering for Collaborative Web Systems: Adapting the Awareness Information to the User’s Context, SAC’05 ACM Symposium on Applied Computing. (2005)
13. Leake, D.,Maguitman, A.,Reichherzer,T.: Exploiting Rich Context: An incremental Approach to Context-Based Web Search. CONTEXT 05 LNCS 2680. Springer Verlag (2005)

Impact of Contextual Information for Hypertext Documents Retrieval

Idir Chibane and Bich-Liên Doan

SUPELEC, Computer Science dpt.
Plateau de Moulon, 3 rue Joliot Curie, 91 192 Gif/Yvette, France
{Idir.Chibane, Bich-Lien.Doan}@supelec.fr

Abstract. Because the notion of context is multi-disciplinary [17], it encompasses lots of issues in Information Retrieval. In this paper, we define the context as the information surrounding one document that is conveyed via the hypertext links. We propose different measures depending on the information chosen to enrich a current document, in order to assess the impact of the contextual information on hypertext documents. Experiments were made over the TREC-9 collections and significant improvement of the precision shows the importance of taking account of the contextual information.

1 Introduction

Since the beginning of the Web, information has become widely-accessed and widely-published. The volume of heterogeneous and distributed information available on the Web has been exponentially and continuously growing. That's why the seeking and selection of relevant information is a very complex and difficult task. Search engines help the final-user in this retrieval task by indexing a part of the Web, but they have very few information concerning the information need of the user. Experiments show that most of user's requests contain 2 or 3 terms. So few numbers of terms often leads to noise and silence in the responses given by search tools. This is a consequence of several reasons that include, among others, the implicit user's information need (for example her intention, the context of the query) and the non use of contextual information of the documents in the indexing phase. Several works on survey attempted to classify different contexts alongside with functional or opposite criteria. For [14], [15] and [16], the context of a document is the information related to the current document that is conveyed through hypertext links, semantic network, or surrounding text. The context is used to enrich the local index of a document with information extracted from its neighbours. Experiments showed that taking account this context provide better precision for certain types of queries.
In this paper, we are particularly interested in the local context of Web resources and we define the **context of Web pages** as the neighbourhood information of pages

which is brought from the hypertext links to all resources directly related to these current pages. In recent years, several information retrieval methods using the information about the link structure have been developed and proved to provide significant enhancement to the performance of Web search in practice. Actually, most of systems based on link structure information combine the content with the popularity measure of the page to rank a query result. Google's PageRank[1] and Keinberg's HITS[2] are two fundamental algorithms employing the hyperlink structure among the Web page. A number of extensions to these two algorithms are also proposed, such as [3][7][8][9][10][11]. All these link analysis algorithms are based on two assumptions: (1) the links convey human endorsement. If there is a link from page A to page B, then we may assume that page A endorses and recommends the content of page B. Thus, the importance of page *A* can, in part, spread to the pages besides B it links to. (2) Pages that are co-cited by a certain page are likely to share the same topic as well as to help retrieval.

The study of the existing systems enabled us to conclude that all ranking functions based on link structure information do not depend on query terms. It decreased significantly the found results precision. Indeed, analysis of the user's behaviours in their research shows that they are not interested in the popular pages, if it does not contain the query terms. In this paper, we first review the related literature on link analysis ranking algorithms. We also present some extension of these algorithms, by defining the context of Web pages as enriched neighbourhood information conveyed through hypertext links and whose importance is computed according to the query terms. Then, we introduce our new link analysis ranking algorithm with the new ranking function and we present experiments on multiple queries, using the proposed algorithm. We also present a comparative of different link analysis ranking algorithms. Last, we discuss results' analysis.

2 Related Work

Various studies suggested that taking account of links between documents increases the quality of information retrieval. PageRank[1] of Google and the HITS[2] of Kleinberg are the basic algorithms using link structure information. Generally, these systems function in two steps. In the first stage, a traditional search engine returns a list of pages in response to user query. In the second stage, these systems take account of the links to rank the documents results. In this section we describe some of previous link analysis ranking algorithms.

PageRank (PR), introduced by L. Page and S.Brin [1], which is part of the ranking algorithm used by google precomputes a rank vector that provides a-priori "importance" estimates for all the pages on the Web. This vector is computed once, offline, and is independent of the search query. At the query time, these importance scores are used in conjunction with query-specific IR scores to rank the query results. PageRank simulates a user navigating randomly in the Web who jumps to a random page with probability (1-d) or follows a random hyperlink (on the current page) with probability d. This process can be modelled with a Markov chain, from where the stationary probability of being in each page can be computed.

Intuitively, this formula means that the PR of a page A depends at the same time on the quality and the number of pages which cites A. For example, the pages pointed by the home page Yahoo! that have a higher PR will be judged of good quality. The PR computations are long and require cleaning the entire Web. Moreover, the results obtaining by Google shows that the algorithm witch compute PageRank value of a page is not completely relevant. The query results do not have sometimes any relationship with research carried out. Because search engines does not take into account semantics, context or user profile. From where, the idea to compute personalized PageRank. Last years, research led to three radically different solutions [6], the modular Pagerank, the BlockRank and the Topic sensitive Pagerank. The three approaches approximate PR with some approximation, although they differ substantially in their computational requirements and in the granularity of personalization achieved.

Considering the Web is a nested structure, the Web graph could be partitioned into blocks according to the different level of Web structure, such as page level, directory level, host level and domain level. We call such constructed Web graph as the *block-based* Web graph, which is shown in Fig.2 (left). Furthermore, the hyperlink at the block level could be divided into two types: Intra-hyperlink and Inter-hyperlink, where inter-hyperlink is the hyperlink that links two Web pages over different blocks while intra-hyperlink is the hyperlink that links two Web pages in the same block. As shown in Fig 2, the dash line represents the intra-hyperlink while the bold line represents the inter-hyperlink. There is several analysis on the block based Web graph. Kamvar et al. [18] propose to utilize the block structure to accelerate the computation of PageRank. Further analysis on the Website block could be seen in [13][15]. And the existed methods about PageRank could be considered as the link analysis based on page level in our approach. However, the intra-link and inter-link are not discriminated to be taken as the same weight although several approaches proposed that the intra-hyperlink in a host maybe less useful in computing the PageRank [7].

In [8], Kleinberg introduced a procedure for identifying web pages that are good hubs or good authorities, in response to a given query. To identify good hubs and authorities, Kleinberg's procedure exploits the graph structure of the web. Each web page is a node and a link from page A to page B is represented by a directed edge from node A to node B. When introducing a query, the procedure first constructs a focused subgraph G, and then computes hubs and authorities scores for each node of G (say N nodes in total). In order to quantify the quality of a page as a hub and an authority, Kleinberg associated every page with a hub and an authority weight. Following the mutual reinforcing relationship between hubs and authorities, Kleinberg defined the hub weight to be the sum of the authority weights of the nodes that are pointed to by the hub, and the authority weight to be the sum of the hub weights that point to this authority.

3 Modeling the context of documents

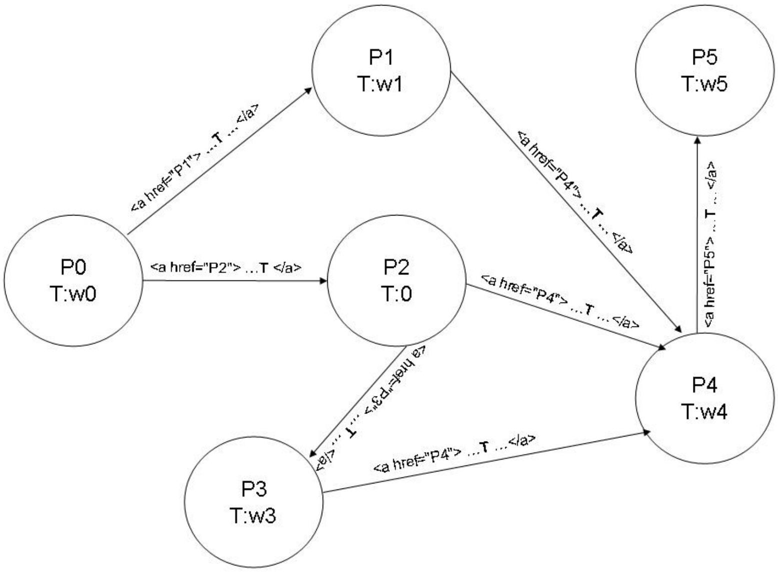
Considering a graph of HTML pages where hypertext links relate source pages to destination pages, and considering the HTML anchor text of a source page that provides information to the destination page. HTML anchors are often surrounded by additionally text that seems to describe the destination page appropriately. The anchor text and the text surrounding an anchor text of a link ("link-context") is used for a variety of tasks associated with Web information retrieval. For example, it may be used by a search engine to rank a page. These tasks can benefit by identifying structural regularities that appear around links and that would constitute a link-context. We describe a framework for conducting such a study. The framework serves as an evaluation platform for comparing various link-context derivation methods. Our focus is on understanding the potential merits of using the zone around the anchor text (link-context), for improving information retrieval. For that, we propose a hyperlink-based term propagation model (HT). The HT model propagates the frequency of query terms in a web page using the context-link information before assigning the relevance weighting algorithms to rank the documents. We consider three types of links: in-link, out-link and in-out-link (bi-directional) (table 1). The HT model can be applied to each type of link by recursively propagating the weight of link-context terms.

Table 1. Applications of the HT model

Weight of link-context	HT propagation function
in-link	$FT^{n+1}(T,D) = FT^{0}(T,D) + \beta * \sum_{(D' \in In(D) \wedge T \in AT(D' \to D))} FT^{n}(T,D')$
out-link	$FT^{n+1}(T,D) = FT^{0}(T,D) + \beta * \sum_{(D' \in Out(D) \wedge T \in AT(D \to D'))} FT^{n}(T,D')$
in-out-link	$FT^{n+1}(T,D) = FT^{0}(T,D) + \beta * \sum_{(D' \in In(D) \cup Out(D)) \wedge (T \in AT(D \to D') \vee T \in AT(D' \to D))} FT^{n}(T,D')$

In the Figure 1, we represent an example of a graph of pages where each node represents a page and each oriented arc from node A to node B represents the link-context to B. Each page contains a set of terms whose weight is calculated by combining the Okapi BM25 score and a term weight propagation using the link-context. It is necessary that these terms appear around the anchor text of links between documents. For example, the weight of the term T in the page P4 is calculated from all the weights of the terms of the pages P0, P1, P2 and P3. The strength of each weight depends on the distance between two documents in terms of links. For example, there are three paths between the page 0 and page 5: P0-P1-P4-P5 and P0-P2-P4-P5 of length 3 and P0-P1-P2-P3-P4-P5 of length 4.

Figure 1. Example of link-context

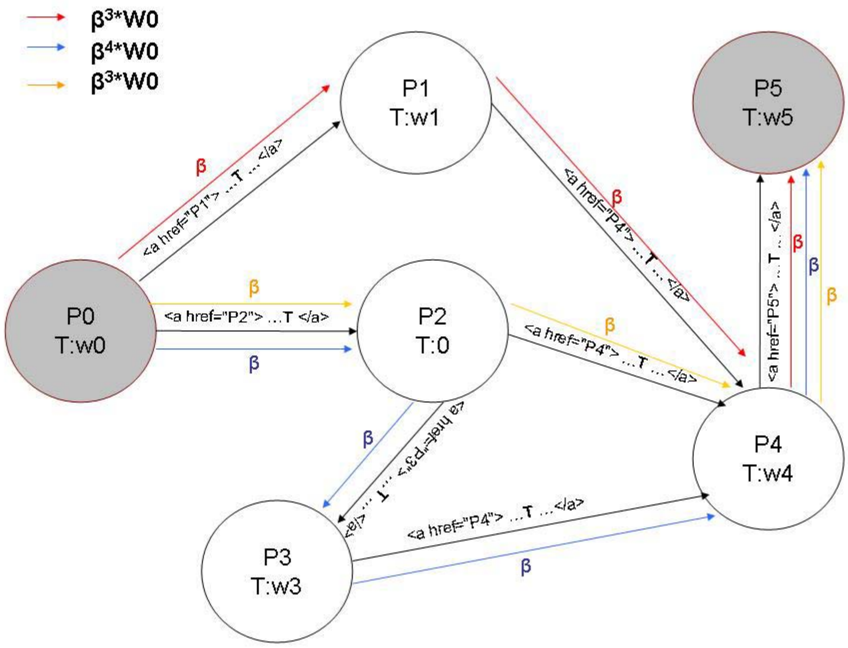


We can easily calculate the weight of the term T in the document D as follow

$$FT^{n+1}(T,D) = FT^{0}(T,D) + \beta * \sum_{(D_i \in In^{-1}(D))} FT^{0}(T,D_i) + \beta^{2} * \sum_{(D_i \in In^{2}(D))} FT^{0}(T,D_i) + \ldots + \beta^{k} * \sum_{(D_i \in In^{k}(D))} FT^{0}(T,D_i)$$

$In^{k}(D)$ represents a set of documents that are at distance K from document D.

Figure 2. Example of contribution of weight term propagation T from P0 to P5



In table 2, we provide an example of successive iterations corresponding to the figure 1, that illustrates our HT algorithm of weight term propagation. We notice that the propagation weight of terms converge towards the red values. The number of iterations is fixed, in order to eliminate the problem of cycles in the graph.

Table 2. Iterations for the HT model

Iteration 1	Iteration 2
$FT^0(P0,T)=W_0$	$FT^1(P0,T)=\mathbf{W_0}$
$FT^0(P1,T)=W_1$	$FT^1(P1,T)= W_1+ \beta^* W_0$
$FT^0(P2,T)=0$	$FT^1(P2,T)= \beta^* W_0$
$FT^0(P3,T)=W_3$	$FT^1(P3,T)=W_3$
$FT^0(P4,T)=W_4$	$FT^1(P4,T)=W_4+ \beta^*(W_1+ W_3)$
$FT^0(P5,T)=W_5$	$FT^1(P5,T)=W_5+ \beta^*W_4$
Iteration 3	**Iteration 4**
$FT^2(P0,T)=\mathbf{W_0}$	$FT^3(P0,T)=\mathbf{W_0}$
$FT^2(P1,T)=\mathbf{W_1+ \beta^*W_0}$	$FT^3(P1,T)=\mathbf{W_1+\beta^*W_0}$
$FT^2(P2,T)=\mathbf{\beta^*W_0}$	$FT^3(P2,T)=\mathbf{\beta^*W_0}$
$FT^2(P3,T)=W_3+ \beta^2{*}W_0$	$FT^3(P3,T)=\mathbf{W_3+\beta^2{*}W_0}$
$FT^2(P4,T)=W_4+\beta^*(W_1+ W_3)+ 2^*\beta^2{*}W_0$	$FT^3(P4,T)=W_4+\beta^*(W_1+W_3)+ (\beta^3+2^*\beta^2)^*W_0$

$FT^2(P5,T)=W_5+\beta*W_4+\beta^2*(W_1+W_3)$	$FT^3(P5,T)=W_5+\beta*W_4+\beta^2*(W_1+W_3)+2*\beta^3*W_0$
$FT^4(P0,T)=\mathbf{W_0}$ $FT^4(P1,T)=\mathbf{W_1+\beta*W_0}$ $FT^4(P2,T)=\mathbf{\beta*W_0}$ $FT^4(P3,T)=\mathbf{W_3+\beta^2*W_0}$ $FT^4(P4,T)=\mathbf{W_4+\beta*(W_1+W_3)+(\beta^3+2*\beta^2)*W_0}$ $FT^4(P5,T)=\mathbf{W_5+\beta*W_4+\beta^2*(W_1+W_3)+(\beta^4+2*\beta^3)*W_0}$	

4 Experiments over TREC-9

In this section we present an experimental evaluation of our proposed algorithm that we compare to a content based model. We chose the WT10g collection. In our experiments, the precision over the 11 standard recall levels which are 0%, 10%, …, 100% is the main evaluation metric, and we also evaluate the main average precision (MAP) and the precision at 5 and 10 documents retrieval (P@5 & P@10).

Figure 3 shows the experimental results on information retrieval using different context-link methods. The first one which is based on the content-only of the page and is presented with the blue line is the baseline algorithm. The others show results by using our HT model of term propagation according to the types of links. The HT model outperforms the content-only baseline, and specifically the HT model of in-link term propagation is better than the others HT models. These results show that the information conveyed by the in-link is the most important to describe a target page.

Figure 3. Results over TREC-9

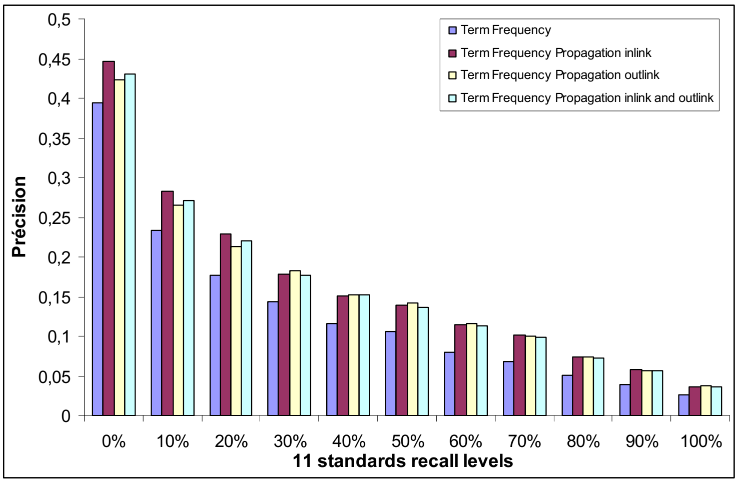


Table 2. Comparisons at MAP, P5 and P10

	TF	TFP_IN	TFP_OUT	TFP_IN_OUT
map	0,1102	0,1416	0,1376	0,1383
P5	0,18	0,22	0,196	0,216
P10	0,148	0,166	0,16	0,16

TF : contents only
TFP_IN : propagation of terms frequency through in-links
TFP_OUT : propagation of terms frequency through.
TFP_IN_OUT : propagation of terms frequency through in-links and out-links.

Table 2 shows that the in-link HT model propagation of terms performs the best result for MAP, P@5 and P@10. For example, the results of in-link HT model propagation achieve 27% for MAP and 22% for P@5.

5 Conclusion

Several algorithms based on link structure to take account of the context of a Web page as an atomic unit of information were developed. But until now, many experiments showed that there is no significant profit compared to the methods based only on content of page. In this paper, we proposed a new method based on link-context using information around the anchor text and the propagation of term weights through the links. We performed experimental evaluations of our system using IR test collection of TREC 9. We conclude that the context of Web pages has a positive impact in the increase of the precision in the top of ranking and in MAP.

We are currently testing our model for expanding queries (relevance feedback) by selecting terms from the surrounding of the anchor text, issued from the co-occurrence matrix between terms of the most relevant documents (we select the top ten relevant documents). Our future work is to test this framework at the semantic blocks level to see the structural effects of blocks on ranking query results. Finally, new measure representing additional semantic information may be explored.

6 References

[1] Brin S. et Page L. (1998), The anatomy of a large-scale hypertextual Web search engine, In Proceeding of WWW7, 1998.
[2] Kleinberg L. (1998), Authoritative sources in a hyperlinked environment, In Proceeding of 9th ACM-SIAM Symposium on Discrete Algorithms, 1998.
[3] Lempel R. et Moran S. (2000), The stochastic approach for link-structure analysis (SALSA) and the TKC effect, In Proceeding of 9th International World Wide Web Conference, 2000.

[4] Savoy J. et Rasolof Y. (2000), Link-Based Retrieval and Distributed Collections, Report of the TREC-9 experiment: Proceedings TREC-9, 2000.
[5] Salton G., Yang C.S. et Yu C.T. (1975), A theory of term importance in automatic text analysis, Journal of the American Society for Information Science and Technology, 1975.
[6] Haveliwala, Taher; Kamvar, Sepandar, Jeh, Glen (2003), An Analytical Comparison of Approaches to Personalizing PageRank, rapport technique, université de Stanford, 2003.
[7] Haveliwala Taher H. (2003), Topic-Sensitive PageRank : A Context-Sensitive Ranking Algorithm for Web Search, Knowledge and Data Engineering, IEEE Transactions on, 2003.
[8] Sepandar D., Kamvar Taher H., Haveliwala Christopher D., Manning Gene H. et Golub (2003), Exploiting the Block Structure of the Web for Computing PageRank, 2003.
[9] Deng Cai; Shipeng Yu; Ji-Rong Wen; Wei-Ying Ma (2004), Block-based Web Search, Microsoft research ASIA, 2004.
[10]Xue-Mei Jiang, Gui-Rong Xue, Wen Guan Song, Hua-Jun Zeng, Zheng Chen, Wei-Ying Ma (2004), Exploiting PageRank at Different Block Level - International Conference on Web Information Systems Engineering, 2004.
[11].Jeh G et Widom. J. Scaling personalized web search. In Proceedings of the Twelfth International World Wide Web Conference, 2003.
[12] Porter M.F. (1980), An algorithm for suffix stripping, 1980.
[13] Ji-Rong Wen, Ruihua Song, Deng Cai, Kailhua Zhu, Shipeng Yu, Shaozhi Ye and Wei-Ying Ma (2004), At the web track of TREC 2003, Microsoft research ASIA, 2004.
[14] Doan, B.L. and Brézillon, P. (2004) How the notion of context can be useful to search tools. Proceedings of the World Conference "E-learn 2004", Washington, DC, USA, Nov. 1-5, 2004
[15] Aguiar, F. Improvement of Web Document Retrieval by the Use of Site's Context Hierarchy. In “Intelligent Exploration of the Web". Springer-Verlag, Heildberg, Germany, 2003.
[16] Mark A. Stairmand. Textual Context Analysis for Information Retrieval. Proceedings of the 20th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR '97, Volume 31 Issue SI. ACM Press. July 1997
[17] M. Bazire et P. Brézillon. "Understanding context before to use it". In 5th International and Interdisciplinary Conference on Modeling and Using Context, Lectures Notes in Artificial Intelligence, Vol 3554, pp. 29--40, Springer-Verlag, 2005.

Discovering Hidden Contextual Factors for Implicit Feedback

Massimo Melucci[1] and Ryen W. White[2]

[1] University of Padova
Department of Information Engineering
Via Gradenigo 6 — 35131 – Padova, Italy
melo@dei.unipd.it
[2] Microsoft Research
One Microsoft Way
Redmond, WA 90852 USA
ryenw@microsoft.com

Abstract. This paper presents a statistical framework based on Principal Component Analysis (PCA) for discovering the contextual factors which most strongly influence user behavior during information-seeking activities. We focus particular attention on explaining how PCA can be used to assist in the discovery of contextual factors. As a demonstration of the utility of PCA, we employ it in an Implicit Relevance Feedback (IRF) algorithm that observes features of user interaction, computes the feature co-variances from a few seen documents, and calculates the eigenvectors of the co-variance matrix to be used as the basis for ranking the unseen documents. This ranking is then compared with the ideal ranking that could be computed if the ratings explicitly given by the user were known. The most effective eigenvector, in terms of impact on retrieval performance, was chosen as representative of each user's intent. Our experiments showed that each aspect of user behavior is influenced by different contextual factors, yet there exist some important features common to most factors. Our findings demonstrate both the effectiveness of the IRF algorithm and the potential value of incorporating multiple sources of interaction evidence in their development. In particular, it was shown that IRF was more effective when the eigenvectors are personalized to each user.

1 Background

Users with vague information needs or limited search experience often require ways to make their queries more precise. Relevance Feedback (RF) [1] provides an effective way of doing this by using relevance information explicitly provided by users. However, despite the promise of RF, users are reluctant to provide explicit feedback, generally because they do not understand its benefits or do not perceive it as being relevant to the attainment of their information goals [2]. As an alternative, implicit RF (IRF) [3] uses features of the interaction between the user and information (e.g., the amount of time a document is in focus in the Web browser or on the desktop, saving, printing, scrolling, click-through), where visited documents to which certain relevance criteria apply are assumed to be relevant. Such contextual features can be mined and used as the basis for relevance criteria in IRF algorithms. These algorithms can suggest query expansion terms, retrieve new search results, or dynamically reorder existing results.

The approach we describe in this paper utilizes user behavior features observed from interaction. Much of the research in this area has focused on the impact on the reliability of interaction features of task and user information [4, 5], click-through [6], session duration and number of result sets returned [7], and document display time [4]. These studies showed that the combination of several implicit features, including display time and the way the user exited from the result page, can predict search result relevance. However, they have also shown that interpreting click-throughs as absolute relevance judgments is sometimes difficult, display times differ significantly according to specific task and user, and that factors such as task, user experience, and stage in the search can affect the potential usefulness of IRF.

As well as developing a better understanding of the accuracy of IRF and the factors that can affect it, work is also ongoing in using this feedback to develop more advanced search systems. Researchers have explored issues such as how behaviors exhibited by users while reading articles from newsgroups could be used as IRF for profile acquisition and filtering [8], to develop a system capable of automatically retrieving documents and recommending URLs to the user based on what the user was typing in a non-search application [9], and to automatically re-rank sentence-based summaries for retrieved documents [10]. To perform these and similar functions IRF has generally been limited to a single behavior such as document display time, editing, or visitation [11–13]. Multiple aspects of user interaction behavior have also been employed [14], but not in the search domain.

In this paper we use multiple aspects of user interaction behavior during search to build models of user interests that can be useful in ranking documents as yet unseen by the user. In the remainder of this paper we describe the approach we adopt and its application for IRF in Section 2, the experiment performed to test its value and its findings in Section 3, and conclude in Section 4.

2 Discovering Hidden Contextual Factors

Information-seeking and retrieval activities are affected by contextual factors that cannot be modeled directly. A *contextual factor* is a variable (e.g., user behavior) that describes one of the possible ways context affects user activity. The *features* are the data observed from user activity. Suppose an observer is trying to understand user behavior when the user is seeking information by measuring various features (e.g., document display time, amount of scrolling). The observer wants to build a model of the user's behavior for modifying the system so as to associate the most relevant documents to that model. Unfortunately, he cannot figure out what is happening because the features appear clouded, sometimes redundant or missing. If a model of user behavior *exists*, then it is hidden behind clouds of noisy data. “Hidden” is related to the latent variables which could not have been observed directly because of the ignorance of the real structure of the information-seeking and retrieval activities. For example the amount of scrolling is likely related to document display time, and therefore one of the two features are probably redundant. What the observer does not know is the degree of redundancy or if some other unobservable variable is governing both features — it may be that this unobservable variable is related to both features so as to make them co-related although they are not when the unobservable is absent. Since these factors are hidden, a mechanism for extracting them is necessary. In this section we present a statistical framework based on Principal Component Analysis (PCA) that can be used to represent these factors in a way that can be leveraged by IR systems for improved retrieval effectiveness.

A statistical framework can discover hidden information from amounts of noisy data [15]. One reason the observed data may be noisy is the absence of (meta-)data about the contextual factors from which the data were observed. In other words, the factors which explain the data are hidden and noise is what makes the data not perfectly explained by the factors. This would mean that a naïve perspective has been taken when observing the data, that is, a perspective for which no noise would exist, and therefore no data about context has been observed. It may be that, if this perspective is changed, noise can be reduced if not removed and the hidden factors governing information-seeking activities can emerge. In the following, we show how the statistical framework can be used for changing perspective and discovering the hidden factors.

When the statistical framework is adopted, the data are naturally represented as vectors and matrices — vector spaces and Linear Algebra is the theoretical framework on which the Vector Space Model for IR (VSM) was proposed in the early Seventies and some recent advances on modeling IR and in particular IR in Context was investigated [16–19, 21]. Thus, when a document or Web page is visited, a feature vector can be associated to it. If k features are observed for each document, the document vectors exist in a k-dimensional vector space. Linear Algebra tells us that every vector in k-dimensional vector space can be represented as a linear combination of k independent *basis* vectors.[1]

In [16, 17], the idea that a document feature vector is the result of a linear combination of basis vectors for representing a document as the result of a combination of hidden contextual factors was presented. Therefore, the discovery of the basis vectors which have generated a document feature vector permits to have a representation of the contextual factors which explain why those features have been observed. With this in mind, suppose a document feature vector has been observed. What is the basis? That is, what are the factors?

This question is important because it points out a tacit assumption which is often overlooked. Indeed, the basis assumed is often the canonical basis. For example, $\{(1,0),(0,1)\}$ is the canonical basis of the two-dimensional vec-

[1] A set of vectors are mutually independent if no vector is a linear combination of the others.

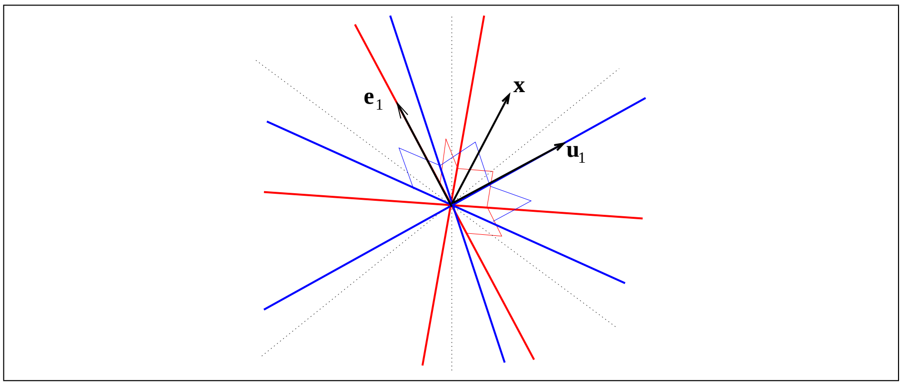


Fig. 1. A vector is generated by infinite dimensions.

tor space and every vector of this space is expressed as linear combination of the canonical vectors. However, nothing prevents us from expressing the same document feature vector as a linear combination of a different basis such as $\left\{\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right), \left(\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}\right)\right\}$ — the vector is the same but its coordinates are different.

In Figure 1 we show how the framework represents a document seen from two perspectives given by two bases. There are two sets of rays — one set of rays is spanned by the basis $E = \{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$, while the other set is spanned by the basis $U = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\}$. Figure 1 depicts how many contextual factors are in the same space. This superposition of factors can naturally be represented by the infinite sets of coordinates which can be defined in the vector space. In the figure, E superposes U. Both E and U can "generate" the same vector $\mathbf{x}$. The myriad of bases model a document or a query from different perspectives and each perspective corresponds to a distinct set of contextual factors. Mathematically, a vector $\mathbf{x}$ is generated by the contextual factors $\{\mathbf{u}_1, \mathbf{u}_2\}$ as $\mathbf{x} = p_1^2\mathbf{u}_1 + p_2^2\mathbf{u}_2 + p_2^3\mathbf{u}_3$ where $\mathbf{u}_i \perp \mathbf{u}_j$, $i \neq j$, $p_1^2 + p_2^2 + p_3^2 = 1$ and $p_i^2 \geq 0$. At the same time, $\mathbf{x} = q_1^2\mathbf{e}_1 + q_2^2\mathbf{e}_2 + q_3^2\mathbf{e}_3$ where $\mathbf{e}_i \perp \mathbf{e}_j$, $i \neq j$, $q_1^2 + q_2^2 + q_3^2 = 1$ and $q_i^2 \geq 0$. An explanation of these expressions is given in [19].

This can be expressed in Linear Algebra using matrices. Let

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_k \end{bmatrix} = \mathbf{I}$$

be the matrix of the canonical basis, which is the starting point of the analysis, that is, the document feature vectors have been observed in this basis. The question is: *Is there another basis which expresses the same vectors and at the same time describe the hidden factors?* The answer is provided by PCA which yields a matrix $\mathbf{C}$ that transforms the feature vectors expressed in the canonical basis into vectors expressed in the new basis.

Mathematically, the change of basis is as follows. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]^\top$ be the $n \times k$ document feature matrix where $\mathbf{x}_i$ is the i-th (row) document feature vector — the column vectors of $\mathbf{X}$ are supposed to have zero means. Let $\mathbf{V} = \mathbf{X}^\top \cdot \mathbf{X}$ be the feature co-variance matrix.[2] PCA yields $\mathbf{C} = \mathbf{U}$ where $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_k]$ is the matrix of the eigenvectors of $\mathbf{V}$. The transformation takes place as $\mathbf{Y} = \mathbf{X} \cdot \mathbf{C}$ so as $\mathbf{y}_i = \mathbf{x}_i \cdot \mathbf{C} = [\mathbf{x}_i \cdot \mathbf{u}_1, \ldots \mathbf{x}_i \cdot \mathbf{u}_k]$. One can recognize that the j-th coefficient of $\mathbf{y}_i$ is the size of the projection of $\mathbf{x}_i$ on to the j-th eigenvector. Because the eigenvectors are mutually orthonormal,

$$\mathbf{x}_i^\top = (\mathbf{x}_i \cdot \mathbf{u}_1)\mathbf{u}_1 + \ldots (\mathbf{x}_i \cdot \mathbf{u}_k)\mathbf{u}_k$$

Therefore, the eigenvectors are the new basis in which the document feature vectors are expressed and are the representation of the contextual factors underlying the generation of the document feature vectors.

Why is PCA so special? In statistics, PCA is used to find the vector of features which best explains the variance of the data. To obtain that, PCA computes the vector which minimizes a function of the co-variance matrix. This vector is the principal eigenvector[3] of the co-variance matrix. The other (orthogonal) eigenvectors capture the residual variance and should represent noise. In our context, the use of PCA and then of a minimum variance-based criterion has the advantage of explaining most of the variance with a small number of eigenvectors. In this way, a few factors can be used for explaining, for example, user behavior. It is useful noting that there are other methods than PCA for representing factors — these methods are classified as decompositions [20].

As the eigenvectors of the co-variance matrix are a representation of the hidden factors, it is natural to be curious about the degree to which a document is affected by a factor. Therefore, the focus is on how the eigenvectors are used for *ranking* documents. If the objects are described by the $\mathbf{x}_i$'s, ranking in context reorders the vectors by the square of the projection between them and the eigenvectors $\mathbf{u}_j$'s which describe the contextual factors. Therefore, the ranking function is

$$|\mathbf{x}_i \cdot \mathbf{u}_j|^2 \tag{1}$$

where $|\mathbf{x}_i| = \sqrt{\sum_j x_{ij}^2} = 1$.

It is interesting to note that the formula resembles the inner product used in the VSM for IR and that PCA was already used in Latent Semantic Analysis for extracting hidden concepts from documents. However, the details of this ranking function can be uncovered and an explanation of why it is proposed can be obtained as reported in [19, 21].

In the following, we describe how to implement an IRF algorithm that captures the contextual factors. As an example, suppose the following six feature (column) vectors have been observed after seeing six (row) documents:

[2] When using PCA, co-variance matrix is suggested.

[3] The principal eigenvector is associated to the largest eigenvalue, which is a measure of the variance explained.

$$\mathbf{X} = \begin{bmatrix} -1.17 & -2.17 & -3.17 & 0.50 & 0.67 & 1.33 \\ -0.17 & -2.17 & 2.83 & 0.50 & -0.33 & 0.33 \\ -0.17 & -2.17 & 0.83 & -0.50 & -1.33 & -0.67 \\ 0.83 & 1.83 & 1.83 & -0.50 & 1.67 & 1.33 \\ 1.83 & -1.17 & -3.17 & -0.50 & -0.33 & -0.67 \\ -1.17 & 5.83 & 0.83 & 0.50 & -0.33 & -1.67 \end{bmatrix}$$

where the columns corresponds to, say, (1) display time, (2) scrolling, (3) saving, (4) bookmarking, (5) access frequency and (6) Web-page depth[4], respectively — all of these values may refer, for example, to time or frequencies, and can be seen as features of user behavior.[5] The following eigenvectors are then computed:

$$\mathbf{U} = \begin{bmatrix} -0.09 & 0.02 & 0.08 & -0.91 & -0.18 & -0.34 \\ 0.91 & 0.37 & 0.10 & -0.05 & -0.17 & 0.04 \\ 0.38 & -0.92 & -0.01 & -0.06 & 0.04 & -0.02 \\ 0.03 & 0.00 & -0.01 & 0.37 & -0.15 & -0.92 \\ 0.04 & 0.05 & 0.69 & -0.03 & 0.71 & -0.13 \\ -0.15 & -0.11 & 0.71 & 0.15 & -0.64 & 0.15 \end{bmatrix}$$

The values of an eigenvector are scalars between -1 and $+1$; the further a value is from 0 the more important it is. In this circumstance, "important" means that the feature to which the value corresponds is a significant descriptor of the contextual factor represented by the eigenvector. The value can be likened to an index term weight. As the values may be negative, the sign can express the contrast between features and then the presence of subgroups of features in the same contextual factor. For example, the first eigenvector, $\mathbf{u}_1$, tells that saving and bookmarking are least important, while the most important feature is scrolling.

Let $\mathbf{u}_j$ be one of these eigenvectors and $\mathbf{x}_i$ be an unseen document. The function of the distance between the document vector and the subspace spanned by the eigenvector is then used as a measure of the distance between the document and the contextual factor. Therefore, $\mathbf{x}_i \cdot \mathbf{u}_j$ is computed. If the unseen document vector is, say, $\mathbf{x}_i = (0.71, 0, 0, 0, 0.71, 0)$, then the distance is 0.03.

In the next section we describe an experiment that compared an IRF algorithm based on PCA that represents each feature separately with a comparator algorithm that uses a single centroid of all features.

3 Implicit Feedback Experiments

The aim of the experiment was to assess the retrieval effectiveness of an IRF algorithm that used the features of user behavior as feedback and translated this feedback into document rankings computed by Equation 1.

[4] The depth of a Web page is the number of links from the root of the Web site to the Web page itself.

[5] This example is inspired by the data set used in D. Kelly. *Understanding Implicit Feedback And Document Preference: A Naturalistic User Study.* PhD thesis, Rutgers, The State University of New Jersey, 2004.

3.1 Methodology

The interaction logs of real subjects were used to simulate a user who accesses a series of Web pages, spends time to read them, scrolls the browser window, moves the mouse and presses keyboard keys. The IRF algorithms under investigation are assumed to be part of a system that monitors user behavior and uses these interaction data as a source of IRF to retrieve and order the unseen documents. When the subject is known, the system records the data by user and then retrieves and ranks the unseen documents for the given user. The details of the simulation are as follows (let us name this algorithm EIG since it is based on the eigenvectors of a feature co-variance matrix):

1. The features of n documents seen by the user are observed and used for computing a representation of context by computing the contextual factors as follows:
   (a) the feature co-variance matrix is computed,
   (b) the eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_k$ are extracted from the co-variance matrix — an eigenvector represents a contextual factor.
2. The documents unseen by the subject are ranked by Equation 1 for each eigenvector $\mathbf{u}_i$.

Multi-level usefulness scores assigned to documents by users have been used as ground truth information for evaluating this IRF algorithm and has not been used for computing the eigenvectors. Normalized Discounted Cumulative Gain (NDCG) [22] was used as a measure of retrieval effectiveness that was able to handle usefulness scores ranging in a non-binary scale.[6] NDCG is a performance metric that is able to make better use of multi-level judgments than precision, which generally must use binary relevance values. NDCG is a measure of distance between two rankings — the ranking produced by an experiment and the best ranking the experiment might produce.

For comparison purposes, the unique centroid vector of the cluster of n vectors of the documents seen by the user was computed. The inner product between the centroid vector and the unseen document vectors is then computed for ranking the unseen documents. Note that no document clustering is performed. Let us name this algorithm CTR. CTR was chosen because it exploits the same data used by EIG but aggregates all interaction feature vectors into a single factor, allowing us to determine the value of utilizing multiple factors, as permitted by EIG.

3.2 Document Features

The data set used in this experiment was gathered during the investigation of the Curious Browser reported in [23]. The set collects the data about 2,127 documents seen by 77 subjects and has information about the actions performed by the subjects whilst conducting self-determined Web browsing tasks, that is,

[6] The discount factor was 2.

without predefined tasks assigned by the experimenter. The following document features of the data set were used in our study: the time spent on a page (`page`), the time spent for horizontal scrolling (`hscroll`), the time spent for vertical scrolling (`vscroll`), the number of scroll events (`#scroll`), the time spent for moving the mouse (`mouse`), the number of the mouse clicks (`#mouse`), the number of times hitting the up arrow key (`#upkey`), the number of times hitting the down arrow key (`#downkey`), the time spent holding the up arrow key (`upkey`), the time spent holding the down arrow key (`downkey`), the time spent holding the page up key (`pgup`), the number of times hitting the page up key (`#pgup`), the time spent holding the page down key (`pgdown`), the number of times hitting the page down key (`#pgdown`), the number of slashes of the visited URL (`urldepth`).[7]

In addition to these features, we also have explicit multi-level ratings assigned by participants based on their own assessment on the usefulness of the document for their browsing activity. These ratings could then be used in the assessment of algorithm performance in our study.

In the next section we present the findings of our study.

3.3 Results

The experiments sought to compare the two IRF algorithms and determine whether there was a contextual factor which orders the unseen documents more effectively than other factors. To this end, the comparison with CTR would allow us to determine whether this "special" eigenvector exists, since CTR computes a single centroid vector. To be precise, the question: "Is there an eigenvector for which EIG "beats" CTR?" will be answered. This "special" eigenvector would allows us to personalize IRF to each user.

In order to establish the role played by the eigenvectors, an analysis was conducted to compare the effectiveness of CTR with the effectiveness of EIG by varying the eigenvector. That is, one eigenvector was fixed at a time and the documents were ranked using the fixed eigenvector. We did this for each subject. Table 1 reports NDCG of CTR and NDCG of EIG. The values in the table are shown after the user had viewed two documents (i.e., $n = 2$). This value has been chosen because it is small enough for evaluating the capability of the simulated system to perform effectively even if the feedback is limited. The number of unseen and ranked documents was $N - n$ where N is the total number of documents seen by the subject in the data set. The eigenvector which achieved the highest average NDCG of EIG was selected over all the eigenvectors. The table reports the composition of the eigenvector for each subject thus making a clear description of the behavior of each subject when accessing the Web pages.

The results suggests that, an eigenvector for which EIG is more effective than CTR almost always exists. Moreover, `page` (i.e., the time spent on each

[7] `urldepth` was added by the authors and was not provided by the data set. The number of slashes has been used because it is a measure of Web-page quality and is an endorsement of the Web-page when the end user selects it. The number of slashes is also known as URL depth and is used for successfully retrieving entry Web pages, which are often preferred by the users when finding resources [24].

lane) is the most important feature of user behavior for every subject. However, the best eigenvector varies its shape depending on the subject. For example, subject 5's behavior is also determined by `mouse` (i.e., time spent moving the mouse). Moreover, some features tend to contrast others. For example, subject 74 spends long periods of time on pages when they seldom scroll, and vice versa. Although `page` describes a common aspect of the interaction of every user, it was clear that each subject had a slightly different interaction style when seeking information, and more than one aspect of this style is necessary to distinguish between subjects. The presence of `page` means that it is necessary for tailoring retrieval to every user, but it is not sufficient since other features are necessary for maximizing retrieval effectiveness. These results suggest that tailoring eigenvectors to users leads to improved performance over algorithms that do not use such an approach. This finding is important because it justifies the design of IRF algorithms that learn from an individual user's interaction and adapt themselves to that user.

Table 1: The composition of the most effective eigenvector for each subject. Feature subgroups corresponding to negative weighs are italicized.

Subject	NDCG		Best Eigenvector
	EIG	CTR	
1	0.883	0.170	`page` (0.843); `vscroll` (0.527); `mouse` (0.107);
2	0.833	0.573	`page` (0.871); `vscroll` (0.442); `mouse` (0.214);
3	0.930	0.491	`page` (0.997); `mouse` (0.078); `pgdown` (0.015);
4	0.907	0.965	`page` (0.894); `mouse` (0.449);
5	0.767	0.654	`page` (0.971); `mouse` (0.238);
6	0.770	0.929	`page` (0.895); `mouse` (0.446);
8	0.933	0.114	`page` (0.746); `mouse` (0.666);
9	0.844	0.804	`page` (0.999); `mouse` (0.023);
10	0.822	0.951	`page` (0.850); *mouse* (-0.53);
11	0.722	0.734	`page` (0.981); `mouse` (0.161); `vscroll` (0.107);
12	0.836	0.741	`page` (0.966); `mouse` (0.253); `vscroll` (0.062);
13	0.916	0.469	`page` (0.957); `mouse` (0.286); `vscroll` (0.051);
14	0.935	0.840	`page` (0.900); `vscroll` (0.386); *mouse* (-0.20);
15	0.873	0.725	`page` (0.995); *mouse* (-0.09);
17	0.738	0.863	`page` (0.915); `mouse` (0.403); `vscroll` (0.025);
19	0.889	0.788	`page` (0.994); *mouse* (-0.04); *pgdown* (-0.10);
20	0.861	0.434	`page` (0.897); `mouse` (0.442);
21	0.658	0.671	`page` (0.827); *mouse* (-0.56);
22	0.868	0.838	`page` (0.967); `mouse` (0.255);
23	0.903	0.501	`page` (0.822); *downkey* (-0.01); *vscroll* (-0.21); *mouse* (-0.53);
24	0.976	0.784	`page` (1.000);

Continued on next page

Table 1 – continued from previous page

Subject	NDCG		Best Eigenvector
	EIG	CTR	
25	0.840	0.827	page (0.991); downkey (0.054); *mouse* (-0.04); *vscroll* (-0.11);
26	0.888	0.777	page (0.950); vscroll (0.264); mouse (0.166);
27	0.751	0.920	page (0.990); mouse (0.138);
28	0.631	0.548	page (0.748); mouse (0.663);
30	0.801	0.519	page (0.999); mouse (0.015);
31	0.938	0.912	page (0.999); mouse (0.015);
32	0.715	0.253	page (0.999); mouse (0.035);
33	0.880	0.626	page (0.925); mouse (0.277); vscroll (0.260);
34	0.849	0.767	page (0.947); mouse (0.322);
35	0.981	0.810	page (0.920); vscroll (0.392); *mouse* (-0.02);
36	0.825	0.411	page (0.859); vscroll (0.473); *mouse* (-0.20);
37	0.892	0.832	page (0.950); mouse (0.304); vscroll (0.048); *pgdown* (-0.01); *upkey* (-0.02); *downkey* (-0.05);
38	0.878	0.930	page (0.994); vscroll (0.093); mouse (0.054); *pgup key* (-0.01); *pgdown* (-0.02);
39	1.000	1.000	page (0.822); downkey (0.383); upkey (0.352); mouse (0.233);
40	0.907	0.848	page (0.913); mouse (0.408);
41	1.000	0.581	page (0.951); mouse (0.310);
42	1.000	0.778	page (0.979); *vscroll* (-0.05); *mouse* (-0.19);
43	0.920	0.898	page (0.999); mouse (0.021);
44	0.981	0.310	page (0.933); mouse (0.355); vscroll (0.056);
45	0.914	0.562	page (0.884); mouse (0.353); vscroll (0.306); pgdown (0.018);
46	0.847	0.630	page (0.997); mouse (0.081);
47	0.893	0.324	page (0.987); *mouse* (-0.16);
48	1.000	1.000	page (0.953); mouse (0.304);
49	1.000	1.000	page (0.953); mouse (0.304);
50	0.961	0.631	page (0.959); mouse (0.283);
51	0.953	0.564	page (0.889); mouse (0.458); hscroll (0.028);
52	0.771	0.605	page (0.900); mouse (0.435);
53	0.892	0.477	page (0.981); mouse (0.100); *vscroll* (-0.17);
54	0.834	0.169	page (0.935); vscroll (0.260); mouse (0.242);
55	0.909	0.587	page (0.883); vscroll (0.052); *mouse* (-0.47);
56	0.946	0.783	page (0.960); vscroll (0.278); mouse (0.032);
57	0.962	0.803	page (0.999); *vscroll* (-0.02);
58	0.887	0.870	page (0.988); *mouse* (-0.15);
59	1.000	1.000	page (0.913); *pgup key* (-0.03); *downkey* (-0.41);
60	0.856	0.789	page (0.891); vscroll (0.323); mouse (0.319);
61	1.000	0.863	page (0.957); mouse (0.252); vscroll (0.140);

Continued on next page

Table 1 – continued from previous page

Subject	NDCG		Best Eigenvector
	EIG	CTR	
63	0.814	0.928	`page` (0.996); `mouse` (0.092);
64	0.920	0.878	`page` (0.932); *`mouse`* (-0.36);
65	1.000	0.995	`page` (0.958); `mouse` (0.248); `vscroll` (0.143);
66	0.901	0.199	`page` (0.868); `vscroll` (0.473); `mouse` (0.152);
67	0.860	0.949	`page` (0.856); `vscroll` (0.472); `mouse` (0.211);
68	0.875	0.944	`page` (0.897); `mouse` (0.443);
69	0.990	0.929	`page` (0.850); `mouse` (0.527);
71	0.958	0.863	`page` (0.976); `vscroll` (0.177); `mouse` (0.127);
72	0.989	0.984	`page` (0.999); `vscroll` (0.042); `mouse` (0.018);
73	0.915	0.884	`page` (0.939); `mouse` (0.345);
74	0.903	0.532	`page` (0.980); *`mouse`* (-0.01); *`vscroll`* (-0.20);
75	0.962	0.224	`page` (0.971); `mouse` (0.199); `vscroll` (0.133);
76	0.713	0.558	`page` (0.995); `mouse` (0.091); `vscroll` (0.030);
77	0.760	0.632	`page` (0.985); `mouse` (0.162); `downkey` (0.051);
Avg.	0.923	0.774	
StDev.	0.088	0.238	

4 Conclusions and Future Work

In this paper a statistical framework that utilizes multiple sources of evidence present in an interaction context has been presented to discover hidden contextual factors that can be used for personalization. The eigenvectors extracted from a feature co-variance matrix observed from interaction are used as representation of the hidden contextual factors. These representations have been compared with an alternative using rich interaction logs (and associated metadata such as relevance judgments) gathered during a user study. Our findings demonstrate the effectiveness of these representations. In particular, it was shown that implicit feedback could be effective when the representation of the contextual factors are personalized to the user. Future work will address the challenge of selecting the best eigenvector automatically.

References

1. G. Salton and M.J. McGill. *Introduction to modern Information Retrieval.* McGraw-Hill, New York, NY, 1983.
2. M. Beaulieu. Experiments with interfaces to support query expansion. *J. of IR*, 53(1):8–19, 1997.
3. D. Kelly and J. Teevan. Implicit feedback for inferring user preference: a bibliography. *SIGIR Forum*, 37(2):18–28, 2003.

4. D. Kelly and N.J. Belkin. Display time as implicit feedback: understanding task effects. In *Proc. of SIGIR*, pages 377–384, New York, NY, USA, 2004. ACM Press.
5. R.W. White, I. Ruthven, and J.M. Jose. A study of factors affecting the utility of implicit relevance feedback. In *Proc. of SIGIR*, pages 35–42, New York, NY, USA, 2005. ACM Press.
6. T. Joachims, L. Granka, B. Pan, H. Hembrooke, and G. Gay. Accurately interpreting clickthrough data as implicit feedback. In *Proc. of SIGIR*, pages 154–161, New York, NY, USA, 2005. ACM Press.
7. S. Fox, K. Karnawat, M. Mydland, S. Dumais, and T. White. Evaluating implicit measures to improve web search. *ACM TOIS*, 23(2):147–168, 2005.
8. M. Morita and Y. Shinoda. Information filtering based on user behavior analysis and best match text retrieval. In *Proc. of SIGIR*, pages 272–281, New York, NY, USA, 1994. Springer-Verlag New York, Inc.
9. J. Budzik and K. Hammond. Watson: Anticipating and contextualizing information needs. In *Meeting of the ASIS*, pages 727–740, Medford, NJ, 1999.
10. R.W. White, I. Ruthven, and J.M. Jose. Finding relevant documents using top ranking sentences: an evaluation of two alternative schemes. In *Proc. of SIGIR*, pages 57–64, New York, NY, USA, 2002. ACM Press.
11. J. Kim, D.W. Oard, and K. Romanik. Using implicit feedback for user modeling in internet and intranet searching. Technical report, College of Library and Information Services, University of Maryland at College Park, 2000.
12. H. Lieberman. Letizia: An agent that assists web browsing. In Chris S. Mellish, editor, *Proc. of IJCAI*, pages 924–929, Montreal, Quebec, Canada, 1995. Morgan Kaufmann publishers Inc.: San Mateo, CA, USA.
13. T. Joachims, D. Freitag, and T. Mitchell. Webwatcher: A tour guide for the world wide web. In *Proc. of IJCAI*, pages 770–777, San Francisco, CA, 1997. Morgan Kaufmann.
14. E. Horvitz, J. Breese, D. Heckerman, D. Hovel, and K. Rommelse. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. In *Proc. of the Conf. on Uncertainty in Artificial Intelligence*, pages 256–265, Madison, WI, July 1998.
15. U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, editors. *Advances in knowledge discovering and data mining*. The MIT Press, 1996.
16. M. Melucci. Can vector space bases model context? In *Proc. of Context-based IR workshop*, Paris, France, July 2005. CEUR.
17. M. Melucci. Context modeling and discovery using vector space bases. In *Proc. of CIKM*, pages 808–815, Bremen, Germany, November 2005. ACM Press.
18. M. Melucci. Ranking in context using vector spaces. In *Proc. of CIKM*, pages 477–478, Arlington, VA, USA, November 2006. ACM Press.
19. M. Melucci. Exploring a mechanics for context-aware information retrieval. In *Proc. of the AAAI Spring Symposium on Quantum Interaction*, Stanford, CA, USA, March 2007. AAAI Press.
20. R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, 1985.
21. C.J. van Rijsbergen. *The Geometry of Information Retrieval*. Cambridge University Press, UK, 2004.
22. K. Jarvëlin and J. Kekäläinen. Cumulated gain-based evaluation of IR techniques. *ACM TOIS*, 20(4):422–446, 2002.
23. M. Claypool, P. Le, M. Waseda, and D. Brown. Implicit interest indicators. In *Proc. of IUI*, pages 33–40, 2001.
24. W. Kraaij, T. Westerveld, and D. Hiemstra. The importance of prior probabilities for entry page search. In *Proc. of SIGIR*, pages 27–34, Tampere, Finland, 2002.

An Ostensive Browsing and Searching on the Web*

Hideo Joho, Robert D. Birbeck, and Joemon M. Jose

{hideo,birbecrd,jj}@dcs.gla.ac.uk
Department of Computing Science, University of Glasgow
17 Lilybank Gardens, Glasgow G12 8QQ, UK.

Abstract. The ostensive model assumes that a user's information need is dynamic and developing, thus, a recently accessed object can be seen as more indicative to the current information need. The model has been proved to be effective in image retrieval. This paper investigates the effectiveness of an ostensive model applied to web retrieval, where query-biased sentences are used to implicitly capture an underlying information need and to support a user's browsing of search results. Our study suggests that the sentence-based approach to an ostensive browsing is promising to facilitate an effective exploration of search results.

1 Introduction

Relevance feedback is one of the critical components in information retrieval (IR) systems. Leveraging a searcher's feedback to improve retrieval effectiveness is a form of system's adaptation to an underlying information need. A criticism of the existing relevance feedback models such as [1] is that they often assume that the underlying information need is static during the search session. Bates [2] and Kuhlthau [3] argue that this does not always represent the searching behaviour of real searchers. They suggest that information needs and search goals are often dynamic and developing during the search. In addition, Pharo and Järvelin [4] suggest that the searching behaviour can be irrational when the searchers face a complex problem. Several models have been proposed by researchers, where the dynamic nature of information needs was taken into account in one way or another [5–8]. Of those, the *ostensive model* (OM) proposed by Campbell and Van Rijsbergen [5] is particularly interesting because it offers a simple but effective way of capturing the developing information need for relevance feedback. The OM has been applied to image retrieval [9, 10]. The model's success in image retrieval appears, partly, to be due to the representation of information objects (e.g., thumbnail image) used in the search result presentation. The representation of objects is important in the OM since it is used by the searcher to interact with the search interface, and since it is used by the system to capture relevance feedback implicitly.

In this paper, we present an application of the ostensive model in Web retrieval, where the top ranking sentences (TRS) [11] are used as the primary representation of information objects for the browsing of search results. There are several motivations for using TRS in our application. First, TRS is a query-biased summary of a document [12], thus, it can be a promising representation for an application of the adaptive models

* This work was supported by EPSRC (Ref: EP/C004108/1).

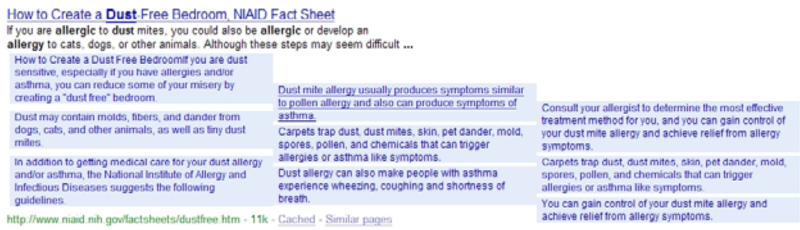


Fig. 1. A screenshot of an ostensive browsing interface

such as the OM. Second, the generation of effective TRS has been established by a series of studies [11, 13–15]. In the existing studies, however, the TRS was presented in a static manner. In this study, the sentences were dynamically ranked by an ostensive model to help searchers find relevant information. The rest of this paper is structured as follows. Section 2 presents the interface of a sentence-based ostensive browsing. Section 3 describes the experimental design of our user study. Section 4 presents the results of the evaluation. Finally, Section 5 discusses our findings and future work.

2 Sentence-based ostensive browsing

Our approach to an ostensive browsing was based on query-biased sentences [12]. For each record of the URLs retrieved by Google, up to three sentences were extracted from the document using a version of the software originally developed by [13]. The software extracted candidate sentences from retrieved documents and ranked them based on a mixture of factors such as the frequency of query terms, document location, and HTML tags. In our interface, the sentences were then appended to the search result, as shown in Fig.1. In the interface, the words from click-through sentences were used to implicitly capture a user's underlying information need. More specifically, when a set of URLs were retrieved in response to a query, the sentences were extracted from the URLs, and content-bearing words were stored in a document-term matrix. The words were given an initial weight based on *TF*IDF* within the set of all top ranking sentences (as opposed to a full-text). When a sentence was accessed in the result, the weight of the words that appeared in the sentence was updated. A new set of sentences were then ranked by the current weight of words and presented to a user. The weight of words was consistently updated as the user interacted with the sentences. A higher weight was given to the words that occurred in a more recently accessed sentence. More specifically, the initial weight was updated by a linear combination with the sum of *ostensive relevance value* [5], defined as $\frac{1}{2^k}$, where k was the distance from the latest interaction. While a more sophisticated function can be used to update the weight [14], we decided to keep it simple since it was not our aim to investigate an optimal ostensive function.

The effectiveness of TRS has been studied in a series of experiments conducted by White, et al. [11, 13–15]. Compared to their system, our interface was intentionally designed to be a simple extension of an existing search engine's result presentation. However, our interface enabled users to browse the retrieved documents via an ostensive presentation of TRS. The main objective of our study is to investigate the effects of a sentence-based ostensive browsing devised for the effective exploration of retrieved documents in a user's information searching behaviour. The next section describes our experimental design to address the research objective.

3 Experiment

A repeated measures within-subject design was used for our experiment, where the independent variables were the system and subject group (see below). The experiment contained a range of dependent variables due to our holistic approach to user-centred evaluation in IIR. Yet, they were largely grouped into participants' browsing of search results, query re/formulation, and their overall task performance. The dependent variables were measured by the post-search questionnaires as well as user interactions with the interfaces recorded by the system. This section presents the details of our experimental design.

Participants A total of 24 participants were recruited for our experiment. The recruitment was carried out by our call for participation distributed to the mailing lists of the University of Glasgow and in a subsequent word-of-mouth fashion. Participants were divided into two groups (twelve each) based on their background. The first group consisted of the undergraduate and postgraduate students in Computer Science (CS) fields who tended to have more search experience than the second group. The second group consisted of the people from various backgrounds (but not CS) who tended to have less search experience than the first group. In this paper, the first group is called *More Experienced* group and denoted as G_1 while the second group is called *Less Experienced* group and denoted as G_2. The entry questionnaire established that the age of our participants ranged from 19 to 50 with an average of 27.8. The average age of the More Experienced and Less Experienced Group was 21.1 and 34.5, respectively. The More Experienced group had on average 7.9 years of search experience (standard deviation: $\sigma = 1.4$) while the Less Experienced group had on average 4.4 years of search experience ($\sigma = 2.0$).

Systems Three systems were devised for our experiment. All systems presented the 10 retrieved records per result page. The first system (System 1, denoted as S_1) was a control system where up to three TRS were appended to the existing document surrogate (title, snippet, url, size, etc.) of individual retrieved records. The presentation of TRS in System 1 was static and no further browsing was available. The second system (System 2, or S_2) was the same as System 1 except that the ostensive presentation of TRS was implemented as discussed in Section 2. When a user *hovered* the mouse pointer on a TRS of retrieved records, three new TRS were extracted from other retrieved records and presented in a cascading menu style. After some informal experimentation on the visualisation, we decided to present up to three levels of menus since it appeared to provide reasonable readability of TRS without cluttering the screen. A more detail measure of appropriate levels for the TRS presentation is beyond the scope of this experiment. When a TRS was *clicked* from the cascading menu, a new window was opened to show the contents of the page where the TRS was extracted. The top 30 retrieved URLs were used to extract and rank TRS for the ostensive browsing. The third system (System 3, or S_3) was the same as System 2 except that query terms were suggested based on user's browsing of TRS. The words appeared in the browsed TRS were recorded and ranked by the OM function. The top six words[1] (except stopwords) were suggested to user by updating the query box in the system interface. We did not include an interface that had no TRS in our experiment because past work (e.g., [13]) has already demonstrated the benefits of TRS compared to such an interface.

Tasks Participants were asked to carry out three search tasks in the experiment. One of our research interests was to evaluate the effectiveness of the proposed interfaces based on a range of search tasks. The tasks were designed based on the simulated work task situation framework [16]. The framework described a task as a form of short scenario. The scenario explained the contexts and motivation of the search with sufficient information about the relevance of pages. An overview of the tasks used in our experiment is as follows.

Task 1: Background search task. This task asked participants to find general background information on a topic. In our experiment, participants were asked to find the pages which provide information about the recent change of student populations.

Task 2: Decision-making task. This task asked participants to make a decision about a topic. In our experiment, participants were asked to find the best Hi-Fi speakers available in a target price. Participants were encouraged to compare the speakers' details in the decision making process. Task 1 and 2 were based on the descriptions originally proposed by [15].

Task 3: Many items task. This task asked participants to find as many items as they feel necessary about a certain topic. In this experiment, the task involved finding out interesting things to do at the city of Kyoto in Japan for a free weekend there. This task was a variant of aspectual search devised in the Interactive Track of TREC [17].

Procedure The user study was carried out in the following manner. At arrival time participants were asked to read an information sheet which described an overview of the experiment and guideline for the participation. Upon the agreement of participation, participants were asked to fill in an entry questionnaire to indicate their background information. Then they were presented with a training topic and explained the nature of simulated-work task. They were given approximately 10 minutes to familiarise with the search interfaces and task activity. During the training session, the three systems were introduced to participants and questions regarding the interface and tasks were answered. During the tasks, participants were asked to bookmark the pages when relevant information was found. However, no explicit instruction was given to participants regarding the number of bookmarks required to complete the tasks. All participants have used the bookmarking function of web browsers in the past and they did not express any difficulty of bookmarking during the experiment. Participants were given 15 minutes to complete a task, but were allowed to end it when they felt they had completed

[1] This size was selected based on a study of a TRS-based system [13].

the tasks. After the first task was completed, participants were asked to fill in a post-search questionnaire to provide subjective assessments about their search. A new task was then given to them and the change of system was informed. The same procedure was repeated three times. Each participant carried out all three tasks using a different order of the three systems. To reduce the bias of system, participants were systematically assigned to one of the following orders of the system: S_1-S_2-S_3, S_1-S_3-S_2, S_2-S_1-S_3, S_2-S_3-S_1, S_3-S_1-S_2, and S_3-S_2-S_1. Since the type and domain of search tasks used in our experiment were different, the order of tasks remained consistent across participants. When the three tasks were completed, participants were asked to fill in an exit questionnaire to indicate their overall preference of system, followed by an open-ended interview to capture their feedback and comments about the result presentation and experiment. The whole session tended to take between 1.5 to 2.5 hours. Participants were rewarded with £5 for their participation.

4 Results and analysis

This section presents the experimental results of our study based on 72 searches carried out by 24 participants. The results presented in this section, unless otherwise stated, is the mean value of 12 and 24 searches for G_1/G_2 and G_{1+2}, respectably, across the systems. The standard deviation of the mean values are given in the brackets. As for the statistical tests, we opted for the non-parametric tests due to the lack of the normal distribution assumed in our data set [18]. The Friedman Test was run to establish the statistical significance ($p \leq .05$) of the differences observed among the three systems (S_1, S_2, and S_3). When a difference was found to be significant, the post hoc test (Wilcoxon Signed Ranks Test) was carried out to find a significant pair(s) through the multiple pairwise comparisons of the three systems. To take an appropriate control of Type I errors, the significance level was set to $p \leq .0167$[2] in the post hoc tests, based on the Bonferroni correction [19]. The same procedure was applied to the results based on all participants (denoted as G_{1+2}). Furthermore, the Mann-Whitney U test was used to establish the statistical significance ($p \leq .05$) of the differences observed between the two subject groups (G_1 and G_2).

This section is structured as follows. Firstly, the experimental results that are related to the browsing of search results are presented. Secondly, we present the results regarding participants' query re/formulation process. Finally, participants' perceptions on the search tasks and their overall task performance are presented, followed by their system preference.

4.1 Browsing of search results

All the systems evaluated in this study presented up to three top ranking sentences (TRS) in the individual retrieved records, in addition to the existing surrogate components such as the title, snippet, URL, and file size. The difference between System 1 and System 2/3 was the functionality of the ostensive presentation of TRS, which was

[2] That is .05 divided by 3 pairwise comparisons.

Table 1. Ease of browsing and finding rel docs (Range: 1-7, Lower = Easier)

Ease of browsing search results	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
More Experienced (G_1)	2.1 (0.9)	2.1 (0.8)	1.8 (0.6)
Less Experienced (G_2)	2.5 (1.5)	2.4 (1.6)	1.7 (1.4)
All participants (G_{1+2})	2.3 (1.2)	2.3 (1.3)	**1.7** (1.0)
Ease of identifying relevant docs	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
More Experienced (G_1)	3.3 (1.1)	2.3 (1.3)	2.0 (0.9)
Less Experienced (G_2)	2.5 (1.5)	2.4 (1.6)	**1.4** (0.7)
All participants (G_{1+2})	2.9 (1.4)	2.3 (1.4)	**1.7** (0.8)

designed to facilitate a user's browsing of search results using a set of query-biased sentences. Therefore, this section investigates the effect of the new presentation for the browsing of search results.

Table 1 shows participants' subjective assessment on the ease of browsing the search results during the tasks. Participants were asked to indicate their assessment by the question "*How easy was it to browse the search results and find the relevant information?*". The assessment was captured by a 7 point scale where a low score represented a more positive perception in the analysis. As can be seen, the difference between System 1 (S_1) and 2 (S_2) was small, while System 3 (S_3) tended to have a more positive score than the other two systems. The result seems to be consistent across the subject groups. The Friedman Tests show that the differences are significant in G_{1+2} ($\chi^2(2) = 7.682$, $p = .022$) but not in the individual subject groups. The post hoc tests show that the difference between S_1 and S_3 in G_{1+2} is statistically significant ($Z = -2.412$, $p = .010$). This suggests that, in overall, participants found S_3 easier to browse the search results and find relevant information than S_1. However, since we did not find a significant difference between S_1 and S_2, the query suggestion offered in S_3 appeared to influence their assessment in this question.

Table 1 also shows participants' assessment on the ease of identifying perceived relevant documents. Participants were asked to indicate their assessment by the question "*How easy was it to identify a relevant document from the results presented?*". As such, this question focused on the relevance assessments on the search results. Participants' perceptions were captured in the same manner as the previous question. As can be seen, both S_2 and S_3 tended to have a more positive score than S_1 in the More Experienced group (G_1). In the Less Experienced group (G_2), on the other hand, the difference between S_1 and S_2 was small but S_3 tended to have a more positive score than the other two systems. The Friedman Tests show that the differences are significant in G_2 ($\chi^2(2) = 10.231$, $p = .003$) and G_{1+2} ($\chi^2(2) = 14.381$, $p = .000$). The post hoc tests show that the difference between S_1 and S_3 is significant in G_2 ($Z = -2.410$, $p = .008$) and G_{1+2} ($Z = -3.388$, $p = .000$). Since the p value in G_1 (.053) was close to .05, we also ran the post hoc test in G_1. The difference was found to be significant between S_1 and S_3 ($Z = -2.570$, $p = .006$), but this should be taken as a tentative result. Overall, these results suggest that participants found S_3 easier to identify relevant documents from the search results compared to S_1. The results also suggest that this trend can be more evident for participants in G_2 than G_1.

Table 2. Number of result pages viewed

	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
More Experienced (G_1)	5.5 (2.8)	5.4 (3.9)	4.8 (2.1)
Less Experienced (G_2)	5.6 (3.1)	4.3 (2.5)	3.9 (2.0)
All participants (G_{1+2})	5.5 (2.9)	4.8 (3.3)	4.4 (2.1)

Table 3. Contribution of layout features (Range: 1-7, Lower = Stronger)

	Title	Snippet	TRS	URL	Size	File Type
System 1 (S_1)	2.0 (1.0)	2.5 (1.0)	3.0 (1.7)	4.5 (2.3)	6.0 (1.6)	5.4 (1.7)
System 2 (S_2)	1.8 (0.9)	2.0 (1.2)	**1.9** (1.0)	4.2 (2.1)	6.0 (1.7)	5.8 (1.6)
System 3 (S_3)	1.6 (0.9)	1.8 (1.0)	**1.6** (0.8)	4.5 (2.1)	6.1 (1.6)	6.0 (1.6)

$N = 24$

Table 2 shows the number of result pages viewed by participants during the tasks. In the experiment, all systems displayed 10 records per result page. However, the ostensive presentation offered in System 2 and 3 allowed participants to access the top 30 records through TRS. Therefore, it was anticipated that the number of result pages which participants viewed to complete the tasks should be reduced. This appeared to be the case, as suggested at the bottom row (All participants) of Table 2. In both subject groups, participants tended to view fewer pages in S_2 and S_3 compared to S_1. However, the Friedman Tests show that the differences among the three systems are not significant. Therefore, while participants' perceptions on the browsing and relevance assessments tended to be more positive when the ostensive browsing and query suggestion were offered in the interface, no conclusive evidence was found for their benefit in the reduction of the number of result pages viewed by participants.

We further investigated the contribution of the individual interface features to participants' decisions of visiting URLs from the search results. The features examined were the title, snippet, TRS, URL, size, and file type of retrieved records. Participants were asked to indicate how strongly each feature contributed to their decision of visiting URLs in the search results. Table 3 shows the result of the analysis. An interesting trend was that while the contribution of the URL, size, and file type tended to remain similar across the systems, TRS' contribution appeared to be increased when the ostensive browsing was available in the interface (i.e., S_2 and S_3). The Friedman Test shows that the difference of TRS is significant across the systems ($\chi^2(2) = 21.031, p = .000$). The post hoc tests shows that the difference between S_1 and S_2 ($Z = -3.157$, $p = .001$) and between S_1 and S_3 ($Z = -3.558$, $p = .000$) are significant. This suggests that participants tended to rely more on TRS to access the URLs from the search results when the ostensive browsing was available, compared to the static presentation in S_1. We also noted that participants gave a more positive score to the title and snippet in S_2 and S_3 compared to S_1. The Spearman's correlation coefficients show that the contributions of TRS and snippet are significantly correlated ($rho = .335$, $p = .004$), and so are the contributions of snippet and title ($rho = .627$, $p = .000$). This suggests that the ostensive browsing had an effect of increasing participants' awareness of the other features of document surrogates during the tasks.

Summary: This section has presented the experimental results regarding the browsing and relevance assessments of the search results. The results show that participants often found the systems with the ostensive presentation easier to browse the search results and identify relevant documents. The results also suggest that the ostensive browsing can lead to an increased level of awareness for the other components of document surrogates.

4.2 Query formulation

Formulating an effective query is often a difficult task for searchers [20]. It has been suggested that a variety of information can be used as the source of a searcher's query re/formulation [21]. In our experiment, the expansion terms were suggested in S_3 based on the interaction with the ostensive browsing of TRS. This section presents the experimental results regarding the query re/formulation.

Table 4 shows participants' subjective assessments on the support of formulating queries offered by the interfaces. Participants were asked to indicate their assessment by the question "*Did the interface increase your ability to formulate relevant queries?*". As can be seen, S_3 were given the most positive score among the three systems in both of the subject groups. The Friedman Tests show that the differences among the three systems are significant in G_1, G_2, and G_{1+2} ($G_1 : \chi^2(2) = 7.171$, $p = .027$. $G_2 : \chi^2(2) = 9.829$, $p = .004$. $G_{1+2} : \chi^2(2) = 16.763$, $p = .000$). The post-hoc tests show that the differences between S_1 and S_2 ($Z = -2.574$, $p = .005$) and between S_1 and S_3 ($Z = -2.257$, $p = .011$) are statistically significant in G_{1+2}. This and the relatively close score between S_2 and S_3 suggest that participants tended to find it easier to formulate queries based not only on the term suggestion function offered in S_3, but also on the overall ostensive browsing that were offered in S_2 and S_3. This also indicates that there is room for improving the way in which suggested terms were presented to participants in S_3. We will elaborate this aspect in Section 5.

Table 5 shows the results of participants' query re/formulation process recorded during the tasks. It presents the number of queries submitted to the interface, unique words used during a task, and average query length. Due to the space limit, it only shows the result in G_{1+2} and no significant difference was found between subject groups. The results show that participants tended to submit a fewer number of queries in S_2 and S_3 compared to S_1. However, the number of unique words and average query length in S_3 appeared to be larger/longer than S_1. The Friedman Tests show that the differences among the three systems are not significant for the number of queries, unique words, nor query length. While the difference was not significant, the task breakdown of the results shows that the number of unique words submitted to S_3 was consistently larger

Table 4. Support of formulating queries (Range: 1-7, Lower = Better)

	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
More Experienced (G_1)	4.3 (1.8)	3.1 (1.4)	2.7 (2.0)
Less Experienced (G_2)	2.8 (1.5)	2.2 (0.6)	1.9 (1.7)
All participants (G_{1+2})	3.5 (1.8)	**2.6** (1.1)	**2.3** (1.9)

Table 5. Number of queries, unique words, and query length (G_{1+2} only)

	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
Queries	5.0 (3.0)	4.0 (3.0)	3.8 (2.1)
Unique words	7.2 (3.4)	6.2 (2.9)	8.6 (4.5)
Query length	4.4 (1.8)	3.7 (1.0)	4.8 (2.1)

$N = 24$

than S_1 in all three tasks. On the other hand, the number of queries submitted to S_3 was smaller than S_1 in Task 2 and 3, and similar (4.9 in S_1 vs. 5.0 in S_3) in Task 1. Therefore, the query suggestion offered in S_3 appeared to help participants diversify search vocabulary to complete a task without increasing the number of queries.

Summary: This section has presented the results regarding the query re/formulation performed by participants during the tasks. While the effect of the ostensive presentation was not always evident in the system logs, there was some indication that suggested that the effort of manually formulating queries can be reduced when the ostensive browsing was available in the interface. This was also partly supported by participants' subjective assessment on the interface's support to query re/formulation.

4.3 Task perceptions and performance

We have discussed the effects of the ostensive presentation of TRS on the browsing of search results and query re/formulation process. This section investigates how these effects influence participants' perceptions on the tasks they carried out. The overall task performance is also analysed in relation to the perceptions.

Table 6 shows participants' subjective assessments on the search tasks they carried out (G_{1+2} only). In particular, the perceptions on the satisfaction of the task outcomes and on the complexity of tasks were investigated. For the satisfaction, the question "*How satisfied are you with the results of this search?*" was asked and the answer was captured by a 7-point scale as before (i.e., *Very (1)* to *Not at all (7)*). For the complexity, participants were asked to indicate a degree of agreement with the following statement "*The search task we asked you to perform was: Very Complex (7) to Very Simple (1)*". As can be seen, the overall difference between the three systems regarding the satisfaction of task outcomes appeared to be small. However, the standard deviation indicates that the assessments on S_1 is likely to be more consistent across participants compared to S_2 or S_3. On the other hand, participants appeared to find the tasks less complex when S_2 or S_3 were used compared to S_1. The trend was consistent across the subject groups. However, the standard deviation on S_2/S_3 was again higher than S_1. The task

Table 6. Participants' perceptions on the tasks (G_{1+2} only)

	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
Satisfaction of search outcomes	2.8 (1.2)	2.9 (1.7)	2.9 (2.2)
Task complexity	3.7 (1.3)	3.3 (1.7)	3.4 (2.3)

N = 24; Range: 1-7; Lower = Better (Satisfaction); Lower = Simpler (Complexity).

Table 7. Number of bookmarked pages and task completion time

Number of bookmarked pages	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
More Experienced (G_1)	3.1 (1.2)	2.5 (2.0)	2.5 (1.7)
Less Experienced (G_2)	4.4 (1.7)	4.2 (2.1)	4.6 (2.8)
All participants (G_{1+2})	3.8 (1.6)	3.3 (2.2)	3.5 (2.5)
Time taken to complete the tasks (sec)	**System 1 (S_1)**	**System 2 (S_2)**	**System 3 (S_3)**
More Experienced (G_1)	666 (150)	715 (151)	609 (162)
Less Experienced (G_2)	724 (77)	724 (139)	716 (145)
All participants (G_{1+2})	695 (120)	719 (142)	662 (160)

breakdown of the results show that the perception on the task complexity in S_2 was consistently better than S_1 in all three tasks while the difference was small. S_3 was given a more noticeably positive assessment in Task 2 and 3 compared to S_1, but it was given a noticeably worse assessment on Task 1. However, the Friedman Tests show that the differences among the three systems are not significant in all results. We also looked at an interaction effect between system and task, and no significant effect was found.

Table 7 shows the number of pages bookmarked by participants when perceived relevant information was found. As can be seen, participants in G_1 appeared to bookmark more pages in S_1 than S_2/S_3. In G_2, participants appeared to bookmark a comparable number of pages between S_1 and S_3. However, the Friedman Tests show that the differences among the three systems are not significant. An interesting point was that G_1 tended to bookmark fewer pages than G_2. The Mann-Whitney U Tests show that the difference between the subject groups was significant in all systems. This suggests that More Experienced group tended to complete the tasks with fewer pages than Less Experienced group across the systems. Table 7 also shows the time taken to complete the tasks in seconds. Overall, participants appeared to complete the tasks faster with S_3 compared to S_1 or S_2 in both subject groups. However, the Friedman Tests show that the differences among the three systems are not significant.

Summary: This section has presented the results regarding participants' perceptions on the tasks they carried out, and their overall task performance. In summary, we did not find much evidence which suggested that the ostensive presentation of TRS had an significant effect on participants' perceptions on the search tasks. While the number of pages bookmarked to complete the tasks can be different across the subject groups, no significant difference was found among the systems regarding the overall task performance.

4.4 System preference

At the end of three tasks, participants were asked to indicate the preference of the systems based on the experience of the searches they carried out. The result is shown in Table 8. As can be seen, participants in both subject groups appeared to prefer S_3 most followed by S_2. Friedman Tests show that the differences among the three systems are significant ($G_1 : \chi^2(2) = 13.500, p = .000$. $G_2 : \chi^2(2) = 16.667, p = .000$. $G_{1+2} : \chi^2(2) = 30.083, p = .000$). The post hoc tests show that the difference between S_1 and S_3 is significant in G_1 ($Z = -2.973$, $p = .001$), the differences between all

Table 8. Participants' system preference (Lower = Better)

	System 1 (S_1)	System 2 (S_2)	System 3 (S_3)
More Experienced (G_1)	2.8 (0.6)	2.0 (0.6)	**1.3** (0.5)
Less Experienced (G_2)	2.8 (0.6)	**2.0** (0.0)	**1.2** (0.6)
All participants (G_{1+2})	2.8 (0.6)	**2.0** (0.4)	**1.2** (0.5)

three systems are significant in G_2 ($Z = -2.887$, $p = .003$), and G_{1+2} ($Z \leq -4.090$, $p = .000$). The Mann-Whitney U test show that the differences between the subject groups are not significant in all systems. While we do not exclude a possibility of participants giving a more positive assessment on S_2 and S_3 just because the interfaces were new to them, these results suggest that participants found a case where the functionality provided by S_2 and S_3 was useful during the tasks.

5 Conclusive discussion

This paper presented a sentence-based approach to the ostensive browsing and searching on the web. A user study with 24 participants was carried out to investigate the effectiveness of our approach. The experimental results have the implications on the effects of the ostensive presentation in the searching process. First, the ostensive presentation can facilitate the effective browsing of retrieved documents. Participants often found the system with the ostensive presentation easier to browse the search results and find relevant information, compared to the static presentation of TRS. With the ostensive browsing available in the interface, participants tended to rely more on them to make a decision of which URLs to visit. An interesting effect we found was that the ostensive presentation appeared to increase a level of participants' awareness on other components of document surrogate. Therefore, the interaction design proposed in this work can be an interesting alternative to the existing TRS presentation such as [14].

Second, the ostensive browsing appears to have a positive effect on a user's formulation of effective queries. Participants tended to submit a fewer number of queries to complete the tasks in System 2 and 3 compared to System 1. While the difference was not statistically significant, participants found System 2 and 3 more supportive of their query re/formulation. The close assessment between System 2 and 3 leads us to believe that the active interaction with TRS had a positive effect on participant's query re/formulation process. However, this also suggests that the way in which the suggested terms are presented should be improved. In the current implementation, the query box was updated with the suggested terms when TRS was accessed. Participants sometimes accepted all suggested terms or delete them all to submit a new query. A better control on the selection of suggested terms should be devised for future system.

In conclusion, our study suggests that the sentence-based approach to an ostensive browsing is promising to facilitate an effective exploration of search results, and further investigation should be carried out.

References

1. Salton, G., Buckley, C.: Improving retrieval performance by relevance feedback. Journal of the American Society for Information Science **41**(4) (1990) 288–297
2. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. Online Review **13** (1989) 407–424
3. Kuhlthau, C.C.: Inside the search process: Information seeking from the user's perspective. Journal of the American Society for Information Science **42**(5) (1991) 361–371
4. Pharo, N., Järvelin, K.: "Irrational" searchers and IR-Rational researchers. Journal of the American Society for Information Science and Technology **57**(2) (2006) 222–232
5. Campbell, I., van Rijsbergen, K.: The ostensive model of developing information needs. In: Proceedings of the COLIS 2., Copenhagen, Denmark (1996) 251–268
6. Chalmers, M., Rodden, K., Brodbeck, D.: The order of things: Activity-centred information access. In: Proceedings of the 7th WWW Conference, Brisbane, Australia (1998) 359–367
7. Chi, E.H., Pirolli, P., Chen, K., Pitkow, J.: Using information scent to model user information needs and actions and the web. In: Proceedings of the CHI conference, Seattle, WA, ACM Press (2001) 490–497
8. Das-Neves, F., Fox, E.A., Yu, X.: Connecting topics in document collections with stepping stones and pathways. In: Proceedings of the 14th CIKM, Bremen, Germany, ACM Press (2005) 91–98
9. Campbell, I.: Interactive evaluation of the ostensive model, using a new test-collection of images with multiple relevance assessments. Journal of Information Retrieval **2**(1) (2000) 87–114
10. Urban, J., Jose, J.M., van Rijsbergen, C.J.: An adaptive technique for content-based image retrieval. Multimedia Tools and Applications **31**(1) (2006) 1–28
11. White, R.W., Jose, J.M., Ruthven, I.: A granular approach to web search result presentation. In: Proceedings of the INTERACT 2003, Zürich, Switzerland, IOS Press (2003) 220–227
12. Tombros, A., Sanderson, M.: Advantages of query-biased summaries in information retrieval. In: Proceedings of the 21st SIGIR Conference, Melbourne, Australia, ACM (1998) 2–10
13. White, R.: Implicit Feedback for Interactive Information Retrieval. Phd thesis, Department of Computing Science, University of Glasgow, Glasgow, UK (2004)
14. White, R.W., Jose, J.M., van Rijsbergen, C.J., Ruthven, I.: A simulated study of implicit feedback models. In: Proceedings of the 26th ECIR, Sunderland, UK (2004) 311–326
15. White, R., Jose, J.M., Ruthven, I.: Using top-ranking sentences to facilitate effective information access. Journal of the American Society for Information Science and Technology **56**(10) (2005) 1113–1125
16. Borlund, P.: Experimental components for the evaluation of interactive information retrieval systems. Journal of Documentation **56**(1) (2000) 71–90
17. Over, P.: Trec-7 interactive track report. In Voorheer, E.M., Harman, D., eds.: NIST Special Publication 500-242: The Seventh Text REtrieval Conference (TREC 7), Gaithersburg, MD, NIST (1998) 33–39
18. Hull, D.: Using statistical testing in the evaluation of retrieval experiment. In: Proceedings of the 16th SIGIR Conference, Pittsburgh, PA, ACM (1993) 329–338
19. Siegel, S., Castellan, J.N.: Nonparametric Statistics for the Behavioral Sciences. 2nd edn. McGraw-Hill (1988)
20. Belkin, N.J.: Helping people find what they don't know. Communications of the ACM **43**(8) (2000) 58–61
21. Spink, A.: Term relevance feedback and query expansion: Relation to design. In: Proceedings of the 17th SIGIR Conference, Berlin, Germany, ACM (1994) 81–90

Process-based and context-sensitive information supply in medical care

Oliver Koch[1]

[1] Fraunhofer Institute for Software and Systems Engineering (ISST)
Emil-Figge-Str. 91, 44227 Dortmund, Germany
{oliver.koch@do.isst.fraunhofer.de}

Abstract: The needs-based information supply for medical workplaces is a necessary condition in order to ensure a maximum quality of medical care. A survey conducted by the Fraunhofer ISST showed that physicians complain about information overload with simultaneous qualitative information poverty in medical practice. The consideration of the respective process and work context can make an important contribution to an improvement of the physicians´ information supply. Within this paper first approaches of a differentiated context model and challenges for the future research work are described.

Keywords: Context Modelling, Information Need, Information Logistics, eHealthcare

1 Motivation

In the context of treatment activities physicians rely on patient information, their diseases, adequate treatment methods and guidelines as well as new research results in the medical and pharmaceutical area. They usually get access to information and knowledge by using medical information systems (clinical information systems, laboratory systems, medical practice management software etc.), local databases, e-resources or traditional print media. Basically it can be assumed that, the amount of available information affects the quality of medical decision making and acting positively. This positive effect, however, is confined by human and individual cognitive limitations. If too much information is available, the phenomenon of "information overload" reduces the performance of human actors rapidly [1]. Information overload occurs whenever the information processing requirements of an individual exceed its information processing capacities. The information processing requirements and abilities can be made measurable over the available and/or necessary time to process this information. Consequences of information overload are that important information is not considered when making decisions, that the ability of setting priorities is reduced and that people find it more difficult to recall previously stored information and are generally more confused.
The causes of information overload can be found in three areas. Firstly the already mentioned problem regarding of the time that is available to process information can

be named. If there is not sufficient time, information can not be processed thoroughly. Closely connected to the problem of processing time is the problem regarding the quantity of information. Available time and information processing capacity affect each other mutually. If there is little time available, only a small quantity of information can be processed. If plenty of time is available, a large information capacity can be used. The characteristics of the information represent a third problem area. If the information is for example very complex, has a high novelty level or is ambiguous, then the data processing capacities are affected negatively.

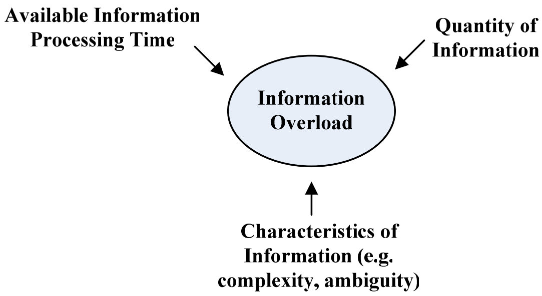


Figure 1: Main Causes of Information Overload

The work of medical staff in healthcare is - to a considerable degree - a highly information-intensive job. Anamnesis data, findings, radiographs, current medication, discharge letters, but also current research results from the clinical research, technical and specialist literature or guidelines must be considered within medical decision making and medical acting. At the same time ever more patient and non-patient-referred information are made available over the Internet and new telematic services (as for example electronic patient records). Therefore the information capacity, which has to be processed by physicians, is very high (problem area 1). If one regards the characteristics of medical information, then these are often complex and ambiguous. The processing of such information makes great demands on the information processing skills of physicians (problem area 3). The fact that sufficient allocations of time are missing in the health service for the processing of information is particularly significant (problem area 2).

In the years 1996 – 1999 the Eurocommunication Study was conducted in six different European countries [2], [3].[1] The differences in physician-patient communication were examined with general practitioners in the six countries. Altogether 190 general practitioners and 2825 patients were included into the study. Concerning the average duration of a doctor-patient consultation the study brought the result that on average in the six examined countries 10.7 minutes are available. In Germany and Great Britain the average consultation duration is even only about 7.6 and accordingly 9.4 minutes.

Within these few minutes the general practitioner must examine and talk to the patient, update electronic patient records and gather additional information, e.g.

[1] Netherlands, Great Britain, Belgium, Germany, Spain and Switzerland.

specialist / technical literature, guidelines, experts contact data. Thus the risk of information overload occurring in an actual treatment situation is particularly high for the physician. This leads to the consequence that often important additional information, e.g. new treatment and therapy methods, current indication-specific guidelines or information gathered from consultations with other experts is not used.

Conceptual and technical mechanisms are missing to integrate such additional information context-sensitive and according to the physicians' information need and also make them available in the physicians` primary systems (clinical information systems or medical practice management software). Thus there is a need for information logistic research, which addresses the following requirements:

1. The physicians' workflow should be interrupted by information retrieval and utilization as less as possible (time factor)
2. The information supply should be reduced to the quantity of information which is necessary and useful in a concrete situation (factor quantity of information).
3. Only such information should be provided that physicians can process in a concrete treatment situation with respect to its complexity, novelty etc.

The information supply should therefore take place regarding to the following information logistic paradigm: Delivering of the right information, in the right amount, at the right point of time to the place, where it is needed. The consideration of context information represents a main lever for the implementation of a need-based information supply in the information logistic research. Beside the physicians' context and its working environment as well as the patient / case context, especially the (treatment) process context, for example in terms of clinical pathways, provides useful information to optimize the information logistic.
First results of the research work on process-oriented and context-sensitive information supply in medical treatment especially the underlying sophisticated context model should be outlined in this paper.

2 Results of a questioning about information need in medical practice

Accompanying to the research work on the topic of "model driven and context-sensitive information supply for medical workplaces" and for the recess of the knowledge concerning the information need of physicians the Fraunhofer ISST conducted a written survey on the information need of physicians at their workplace in the neighboring German large cities Essen and Bochum in 2005 in the context of the project „Needs-based supporting of physicians at their workplace by information logistic applications“ [4].[2]

The data acquisition took place as complete survey under all 2.543 physicians of the cities Bochum and Essen. This total number of respondent physicians divides into about 1.500 resident physicians, 1.000 hospital physicians as well as 40 company medical officers. All medical specialist areas were represented in the questioning. With 240 answered questionnaires the total return ratio amounts to 9.4%. Concerning the professional experience and thus the age structure of the answering physicians, for example with younger physicians, no significant emphasis was recognizable. 43 % of the answering physicians have more than 15 years of professional experience.

The dispatched questionnaire consisted of 28 open and 4 closed questions. Questions about the research area ‘information need’ and ‘information seeking’ behaviour were placed in the questionnaire. In the following in short form selected results of the questioning in statement form are presented, which were of interest for the research work:

Statement 1: The acquisition and processing of information are connected for physicians with high expenditure of time. 40% of the physicians indicate that they spend more than 6 h per week with information retrieval. About 30% of the physicians need besides again more than 6 h per week, in order to evaluate the information. Physicians stop the information retrieval in on an average 30% of the cases due to lack of success.

Statement 2: The physicians still expect an increase of the temporal effort for the information retrieval for the future. The time expenditure for the information acquisition is assessed highly till very highly of 68% of the physicians. Clearly over 50% of the physicians expect that the time requirement will in the future still increase.

Statement 3: The majority of the physicians want to have the same information supply with sinking expenditure for the information acquisition procurement. 59 % of the physicians prefer a constant information supply and information quality with less expenditure of time for the acquisition of the information. This preference is more strongly pronounced with resident physicians (62%) than with hospital physicians (56%).

Statement 4: The respective activity in the treatment process determines the physicians’ information need. About 70 % of the interviewed physicians indicate that their information need is high with the reporting of findings and the diagnosis. With the progressing of the treatment process the information need sinks. During the diagnosis (67 % of the answers) and therapy (58 % of the answers) frequently ad hoc and situation-referred information need occurs.

---

[2] The project was accomplished in co-operation with the national Ministry of Health (North Rhine-Westphalia), promoted by the state chancellery of North Rhine-Westphalia and co-financed with funds of the European Union.

Statement 5: The present ICT solutions for information retrieval are rated as too slow and not easy to handle. 62 % of the physicians confirmed the statement that the information acquisition takes to long. 71 % stated that the information supply is too unclear and 67 % called the frequently inaccurate search results as negative point. With a value of 2.5 on a scale of 0 - 3 the problem that the information acquisition takes to much time was called particularly strong.

It can recapitulatory be noted as result of the questioning that information acquisition and processing are important activities of physicians (Statement 1). Both activities are however afflicted - despite the fact that physicians are knowledge workers - in practice with substantial problems. The statement that about 30 % of the information retrieval activities are abandoned because of unsuccessfulness makes this impressively clear. On the basis of different statements the problem of information overload and its causes is mentioned. The factor time is particularly critical (-> Statement 2 and 3). Physicians don't wish any more information, but a temporally more efficient access to information. In doing so, the treatment process and the particular activity in the treatment process, which is the trigger for the specific information need, build up the main context of the information supply (Statement 4). A context is defined below as „any information that can be used to characterize the situation of an entity“. The hitherto available concepts and solutions aren't able to give physicians a quick access to the information needed and make them available context-sensitive and for this reason need-oriented and quickly available (Statement 5).

As a result of the study the retrieval and working context of physicians was identified as a main focus of the further research work in information logistics. The results also indicate that the physicians' information supply can be significantly improved by explicit considering of context information in information retrieval.

3 A context-sensitive information supply for medical workplaces

Since the beginning of the 90's the issue areas of „Context“ and „Context Modelling“ are aspects of research in the field of „Ubiquitous Computing“ and „Mobile Computing“ [5], [6], [7], [8], [9]. Computers within the meaning of Ubiquitous Computing mostly communicate by using mobile ad-hoc networks. Computers and devices in the context of Ubiquitous Computing are equipped with information and communication facilities and „know“, where they are, which other devices and computers are nearby and what happened to them in the past.

The context dimension „location“ plays a major role in the mobile computing research area To know where a person or device is located, is essential for the need-based provision of services and data. Consequently the context dimension „location” is a central element of all context models in the literature of ubiquitous or mobile computing. Additional recurring context dimensions are „time”, „nearby objects and/or persons” as well as in some cases the actual user activity. Other context attributes can be derived from the specifity of the particular application.

The context models of the mobile and ubiquitous computing are in particular in respect to the weighting of the context dimensions only partly transferable to a need-bases information supply in healthcare. Information supply and logistics are partial aspects of information retrieval, i.e. also the corresponding context model must refer to information retrieval and not to ubiquitous computing [10], [11], [12]. The research on the subject of process-oriented and context-sensitive information supply in medical care can't take place regardless of the actual information technology background in the healthcare sector.

Physicians, who need information in the context of medical care, e.g. relating to recommended differential diagnostic activities according to the appropriate medical guideline, are usually users of certain information systems (e.g. clinical information systems, medical practice management software etc.)

Information delivered in the context of information pull or push should make accessible in the actually used information system to avoid media disruption and change of system.

Concerning medical information systems one inevitably comes across the standardization efforts of the HL7 organization. HL7 is an international standard for the exchange of data between computer systems in the health sector. HL7 provides interoperability between Clinical Information Systems (CIS), Medical Practice Management Software, Laboratory Information Systems as well as Medical Accounting Systems and Electronic Medical Record (EMR). Health Level Seven is one of several American National Standards Institute (ANSI) -accredited Standards Developing Organizations (SDOs) operating in the healthcare arena. Beside the creation and setting of data exchange standards in medical environments, the HL7-organization is engaged in creating mark-up standards for clinical documents (for Clinical Document Architecture (CDA)). The Technical Committees of the HL7-organization work on the improvement and evolution of the different HL7-Standards. The idea of information retrieval solutions for the supply of medical staff with additional information is one of the side aspects in the work of the Decision Support Technical Committee.

One important approach to implement a context-sensitive supply with additional information at the physicians' workplace is the HL7-Infobutton [13], [14], [15], [16]. "An infobutton is a point-of-care information retrieval application that automatically generates and sends queries to electronic health information resources (e-resources) using patient data extracted from the electronic medical record and context information that is captured from the interaction between a clinical user and a clinical information system [...]."

An information request (as a HL7 message) to an e-resource or a local database triggered by the HL7 Infobutton includes particular context information. These are basically elementary patient attributes (age, gender), the physicians' actual activity (e.g. patient information review) and main search concept (e.g. lab parameters or diagnosis) including additional qualifying attributes and information about the physicians' role and language.

Section of a XML schema instance of the infobutton standard context parameter

```
<searchParameter>
   <mainSearchCriteria>
      <mainSearchConcept code="363406005" codeSystem=
      "SNOMED- CT" displayName= "colon cancer">
         <originalText>adenocarcinoma of the colon
         </originalText>
      </mainSearchConcept>
      <modifier code="D011379" codeSystem="MeSH"
      displayName="Prognosis"/>
   </mainSearchCriteria>
</searchParameter>
<searchContext>
   <taskContext>
      <task code="11" />
   </taskContext>
   <patientContext>
      <age value="68" />
      <gender code="F" />
   </patientContext>
   <userContext>
      <role code="C11599" />
      <discipline code="C13429" />
      <language code="eng" />
   </userContext>
</searchContext>
```

Compared with the physicians' real context this context information represents only a small excerpt. Capabilities to expand this approach can be identified especially in the following areas:

1. The infobutton context model isn't process-oriented. Implicit or explicit information need of physicians expresses itself in the context of a treatment process, which is if applicable even formalized by a clinical pathway or a medical guideline. It is of great concern that the context of the information need is not only related to a point of time (actual activity), but rather to a period of time (treatment process). Which activities have already been passed (e.g. anamnesis, differential diagnostics)? Which experience does the physician have with this kind of clinical pathway? What are the next activities in the process? The answers to these questions (process context) are important for the interpretation of the information need.
2. Because of the subsumption into the HL7 context the infobutton context model includes naturally only such context elements that are part of the Reference Information Model (RIM) for HL7 messages. In so far key context information outside the Reference Information Model are missing, e.g. process information, user profile (medical specialty, work experience, preferences concerning specific e-ressources), physicians' system environment (PDA/tablet-PC/workstation or CIS/RIS etc.) and so on.

These two aspects are the starting point for the proposed extended context model of the physicians' process-oriented information supply outlined in the following chapter.

4 An Extended Context Model of Information Retrieval for Medical Workplaces

The above named deficit areas (process-orientation and additional context elements) were specifically addressed during the composition of the first context model version. The selection of the context areas and the describing attributes within the context model takes place based on the analysis of the few existing context models of physicians information retrieval, the results of the questioning of physicians in Essen and Bochum conducted by the Fraunhofer ISST as well as own reflections.

In doing so the selected attributes have to meet three key criteria:

1. The context attributes have to be appropriate to satisfy the subjective and especially the objective information need of doctors to a preferably high degree. In what way this criterions can be matched by the context attributes, can eventually be ascertained only on the basis of empirical tests in cooperation with physicians and comparative studies of the success of medical treatment in scenarios where the physicians were supplied with additional information (context-based information retrieval) and where no additional information was provided.

2. Automatic acquisition and collection of context information should be possible to a high degree. I.e., context information should ideally be stored in digital form in the workplace information systems or a separate context storage system (e.g. patients' primary diagnosis, physicians' medical speciality). If the acquisition of the actual context information doesn't arise automatically within the treatment process, a manual acquisition is necessary. This can be needful in the case of using physicians profile data. The context acquisition should be preferably required only one-time.

3. In the context of information retrieval context information can be usefully utilized only if they have a significant influence on the search result list. I.e. it isn't adequate, when the context information slips into the formulation of a search term, but it has to be used during the execution of the search term by the search engine of the e-ressource or the local database. For this purpose a mapping of context information on objects and data fields in the retrieved information base, e.g. within the framework of a search ontology, is required.

Within information retrieval behaviour of physicians one can differentiate between the active pull information access and the passive push information supply. In the context of information pull a physician is usually searching for further information starting from a key search question (based on one or more search terms). This form of active search has been considered in the extended context model by an optional model area, i.e. there is one (or more) key search term, which is the starting point of the information retrieval. Within push information supply a physician is provided with additional information that - related to the actual activity in the context of a treatment process – can contribute to the improvement of medical decision making and acting, which additional information can be useful, is derived from the physicians' context because in this case a key search term is usually missing.

The process-oriented context model of physicians' information retrieval is divided into four areas:

A. patient context
B. physician context
C. process context
D. environment context

At the pull access a fifth area is added, which contains the key search term and additional information related to this term.

The **patient context** covers the physicians' treatment context, in which he is actually situated und from which the information need is arising. Attributes of the patient context are for example the actual primary and secondary diagnosis, relevant findings, the actual medication as well as the patient's age and gender.

Information about the physician, who expresses an information need, is represented by the **physician context**. The doctor's medical specialty and if necessary further specializations, his work experience and role, but also his preferences regarding specific topics of interest or presentation forms of knowledge and information are described in this model. Eventually the physicians' context can be interpreted as a kind of extended user.

The sophisticated **process context** as the third component of the context model represents the main extension in comparison with existing approaches. The embedding of physicians' activities into a medical workflow is of fundamental importance to anticipate which information is needed by a physician in the next workflow step and how the information can be delivered. It is also important to know, which is the actual activity, which are the following activities and more comprehensive process sequences, how often the physician has passed through a specific process type and who are further parties involved in the medical process.

The embedding of a physician into a physical and IT technical environment is captured by the **environment context**. In this context area it is captured, where the physician is located, which workplace system and end-user devices he is actually using, which other persons and devices are nearby and at which point of time the physicians' information need is expressed. The environment context includes the essential attributes, which are also relevant in the context of ubiquitous and mobile computing.

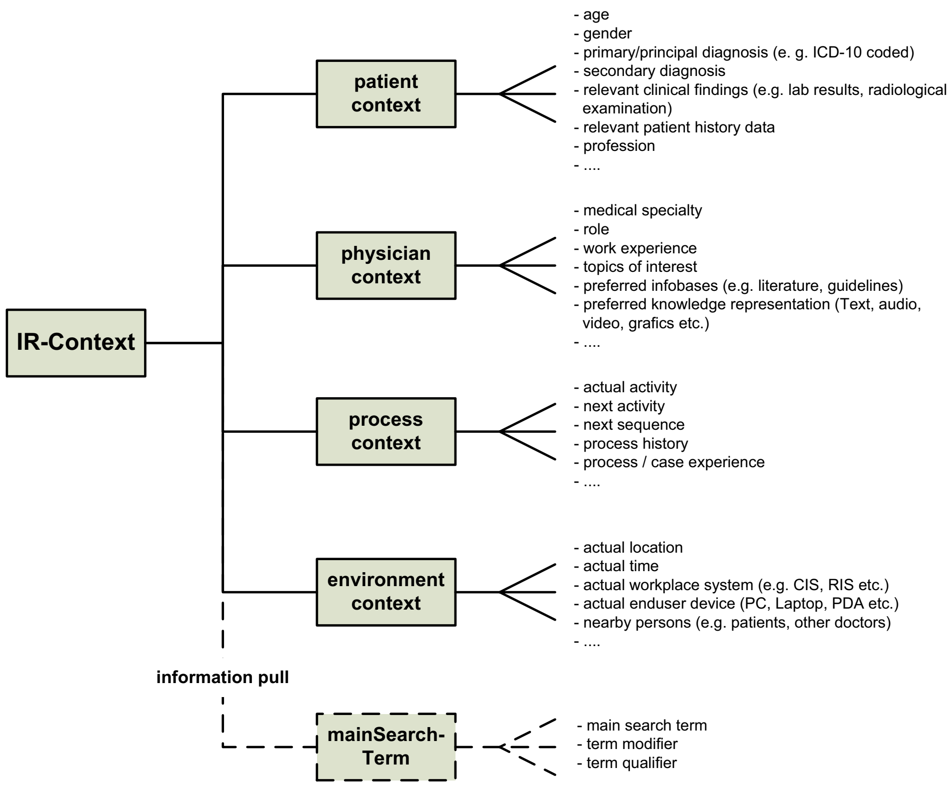


Figure 2: Extended context model of information retrieval at medical workplaces

As previously mentioned the main search term as fifth context area is of high relevance for a physician in the context of active information pull. The main search term can be further specified by additional and qualifying information.

5 Summary and Outlook

Physicians' information supply in medical practice can often be characterised by information poverty with simultaneous information flooding. Due to lack of time, information complexity and information quantity important additional information, which can contribute to the improvement of medical acting and decision making, are not included into treatment processes. Context-based and process-oriented information supply can offer a starting point for the improvement of this problem definition. In the course of this a differentiated context model of the physicians' information need is a key element of information retrieval. An appropriate model was outlined in this paper.

In the future research work this context model and the single context attributes should be further detailed and validated by empirical studies with physicians. For this purpose it is necessary to develop an information logistic prototype, which is based on the extended context model. Starting form a clinical pathway and based on the extended context model physicians should be supplied with additional information, which can be retrieved from external e-Resources (e.g. Pubmed) and internal databases (e.g. clinical information system).

References

1. Eppler, M.J., Mengis, J.: The Concept of Information Overload, MCM Research Paper (2002)
2. van den Brink-Muinen, A., Verhaak P., Bensing, J., Bahrs, O., Deveugele, M., Gask, L., Mead, N., Leiva-Fernandez, F., Messerli, V., Oppizzi, L., Peltenburg, M., Perez, A.: Doctor-patient communication in different European health care systems: relevance and performance from the patients' perspective. In: Family Practice 20 (4) (2003) 478 – 485
3. Deveugele, M., Derese, A., van den Brink-Muinen, A.; Bensing, J., De Maeseneer, J.: Consultation length in general practice: cross sectional study in six European countries. In: British medical journal, 2002 Aug 31, 325 (2002) 472 – 474
4. Reuter, C., Vollmer, G., Koch, O.: Work process-oriented implementation of medical guidelines. In: Shoniregun, C.A.: Proceedings of the International Conference on Information Society (i-Society 2006), August 7-10, Miami (2006) 17-24
5. Dey, A.K., Abowd, G.D.: Towards a better understanding of context and context-awareness. In: Morse, D.R., Dey, A.K. [Eds.]: Proceedings of the CHI 2000 Workshop on The What, Who, Where, When, and How of Context-Awareness, The Hague, Netherlands (2000) 304 – 307
6. Dey, A.K.: Providing Architectural Support for Building Context-Aware Applications. PhD Thesis, Georgia Institute of Technology (2000)
7. Schilit, B., Adams, N., Want, R.: Context-Aware Computing Applications. In: Proceedings of IEEE Workshop on Mobile Computing Systems and Applications, Santa Cruz (2004) 85 – 90
8. Selker, T.; Burleson, W.: Context-aware design and interaction in computer systems. In: IBM Systems Journal 39 (3/4) (2000) 880 – 891
9. Schmidt, A., Beigl, M., Gellersen, H.-W.: There is more to context than location. In: Computers & Graphics Journal, Elsevier 23 (6) (1999) 893-902
10. Gross, T., Klemke, R.: Context Modelling for Information Retrieval - Requirements and Approaches. In: IADIS International Journal on WWW/Internet 1 (1) (2003) 29-42
11. Ingwersen, P., Järvelin, K.: Information retrieval in contexts. In: Ingwersen, P., van Rijsbergen, K., Belkin, N., Larsen, B. [Eds.]: Information Retrieval in Context: Proceedings of the ACM-SIGIR Workshop 2004, Sheffield University (2004) 6-9
12. Jahnke, J., Bychkov, Y., Dahlem, D., Kawasme, L.: Context-Aware Information Services for Health Care. In: Roth-Berghofer, T.; Schulz, S. [Eds.]: Proceedings of the KI-04 Workshop on Modeling and Retrieval of Context, Ulm, September 20-21 (2004)
13. Cimino, J.J., Lia, J., Allena, M., Currieb, L.M., Grahama, M., Janetzkia, V., Leeb, N.-J., Bakken, S., Patela, V.L.: Practical Considerations for Exploiting the World Wide Web to Create Infobuttons. In: Medinfo 11 (2004) 277-281
14. Cimino, J.J., Li, J., Bakken, S., Patel, V. L.: Theoretical, Empirical and Practical Approaches to Resolving the Unmet Information Needs of Clinical Information System

Users. In: Kohane, I.S. [Eds.]: Proceedings of the AMIA 2002 Annual Symposium, American Medical Informatics Association, San Antonio (2002) 170 – 174

15. Cimino, J.J., Meyer, M., Lee, N.-J., Bakken, S.: Using Patient Data to Retrieve Health Knowledge. In: AMIA 2005 Annual Symposium Proceedings (2005) 136 – 140

16. Del Fiol, G., Cimino, J.J., Rocha, R., Maviglia, S., Strasberg, H., Kuhn, T.: Infobutton Communication Standard Proposal, Draft Standard for Trial Use (DSTU), Health Level 7, Date: 2006 Mar 10 (2006)

Improving LSA by expanding the contexts

Nicolas Béchet, Mathieu Roche, and Jacques Chauché

LIRMM - UMR 5506, CNRS, Univ. Montpellier 2,
{nicolas.bechet,mroche,chauche}@lirmm.fr

Abstract. Latent Semantic Analysis is used in many research fields with several applications of classifications. We propose to improve LSA with additional semantic information found with syntactic knowledge.

1 Introduction

In this paper, we use the Latent Semantic Analysis (LSA) approach [1]. LSA is a statistic method applied to high dimension corpora to gather terms (conceptual classification) or contexts (textual classification). The proximity between terms or contexts provided by LSA represents a first step of classification tasks. Our approach, *ExpLSA* (**Exp**ansion of Contexts with **LSA**), consists in expanding the context. This expansion is based on semantic information found with syntactic knowledge. In this paper, we use a Human Resources corpus of PerformanSe company[1] (3784 KB) in French.

For the LSA method, the words that appear in the same context are semantically close. A corpus is represented by a matrix. The lines represent the words and the columns are the different contexts (document, section, sentence, etc). Each cell in the matrix stands for the number of words in a context. Two semantically close words have vectors close (lines of the matrix). The proximity measure is generally defined by the cosine between the vectors.

LSA is based on the Singular Value Decomposition (SVD) theory. $A = [a_{ij}]$ where a_{ij} is the frequency of word i in the context j, breaking down in a product of three matrices USV^T. U and V are orthogonal matrices and S a diagonal matrix. Let us S_k where $k < r$ the matrix built by removing of S the $r - k$ columns which have the smallest singular values. We take U_k and V_k, matrices obtained by removing corresponding columns of U and V matrices. Then, the $U_k S_k V_k^T$ can be considered like an approximation of the version of the original matrix A. Experiments presented in section 4 are applied with a factor $k = 50$, a low value that is more suitable for small corpora.

Before the singular value decomposition, a first step of normalization of original matrix A is applied. This normalization consists in computing a logarithm and an entropy computation on matrix A. This process allows to estimate the weight of words in their contexts. This normalization can also be based on the tf×idf method, a well-known approach in the field of the Information Retrieval

[1] http://www.performanse.fr/

(IR). Let us note that the punctuations and stop words (like "and", "a", "with", etc) are not taken into account to compute LSA.

LSA has many advantages like the languages and domains independence. Nevertheless, an important limit of LSA is based on the size of contexts. Rehder *et al.* showed that the contexts with less than 60 words obtain disappointing results [2].

2 State-of-the-art based on the addition of syntax to LSA

The approaches described in [3, 5] take into account the syntactical knowledge. The approach of [3] uses the Brill's tagger [4] to assign a part-of-speech tag to every word. With this method, LSA considers each word/tag as a single term. This method gives disappointing results. The second approach described in [3] is based on the syntactic analysis in order to segment a text. A syntactic analysis of sentences on three elements (subject, verb, and object) is firstly done. Then, the similarity (cosine) is calculated separately for the three elements (three LSA matrices). The average of the similarities is finally computed. This method gave satisfactory results compared to "traditional LSA".

The approach described in [5] proposes a model called SELSA. It uses part-of-speech tag and a "prefix" label. This one informs about the syntactic type of the words' neighborhood. This approach is close to [3] but SELSA extends this work by generalizing it. A word with a syntactic context specified by its adjacent words is seen as a unit representation of knowledge. SELSA makes less errors than LSA but these errors are more harmful.

In our work, the contexts are represented by sentences. They have a small size giving low results with the LSA method [2]. We propose to use the regularity of some syntactic relations in order to expand the context.

3 Our approach: *ExpLSA*

The final aim consists in automatically gathering terms (conceptual classification) extracted by a system like SYNTEX [6] or EXIT [7]. We propose to gather nominal terms extracted with EXIT from the Human Resources corpus. LSA and *ExpLSA* are the first stage for the conceptual classification task.

The first step of the *ExpLSA* approach identifies the different terms extracted by EXIT. This process consists in representing each term by only one word (for instance, the french term *attitude profondément participative* becomes *noun234* which is the 234th term of a list extracted by EXIT).

After this process, SYGMART parser [8] is applied. This one gives the syntactic relations of each sentence. In our approach, we study Verb-Object relations (Verb_Object, Verb_Preposition_Complement) of our corpus.

The next step of our approach studies semantic proximity between verbs using the Asium measure [9]. With this measure, the verbs are semantically close when they have a lot of common objects. In the next section (section 4), several Asium thresholds based on the similarity values between the verbs will be presented. When the values of the Asium threshold are high, the verbs are close.

The next step proposes to gather common objects (words) of close verbs. Words of the corpus are replaced with all the words of its same group built at the precedent step. For example, our initial lemmatized sentence in French: "*Votre* ***interlocuteur*** *être donc bien inspiré...*" becomes finally: "*Votre* ***(interlocuteur collaborateur)*** *être donc bien inspiré...*". LSA can be applied with the expanded corpus. Very general nouns are not selected to expand context (as "chose" (*thing*), "personne" (*person*), etc).

4 Experiments

In these experiments, we compare similarities given by LSA/*ExpLSA* with a manual expertise. The experts have manually associated terms to 17 concepts. For instance, with our corpus, the expert defined "Relationnel" (*relational*) concept where the term *contact superficiel (superficial contact)* is an instance.

The five most representative terms (the most frequent) which are instances of concepts are used in our experiments. The similarity (cosine) for all representative terms of two concepts is computed. We can verify that the most close pairs of terms given by LSA and *ExpLSA* are instances of the same concept (i.e. these pairs are called *relevant*). In order to compare the results of similarity returned by LSA and *ExpLSA*[2], we propose to calculate the ranking sum of relevant pairs of terms. Then, in our experiments, with the lower sum, we obtain the better results. This evaluation measure is an approach based on ROC curves (Receiver Operating Characteristic) and Area Under these Curves [10]. This feature is mostly used to compare ranking functions [11]. The Area Under ROC Curves is equivalent to calculate the sum of the relevant elements [12].

Pairs of concepts	**LSA**	**ExpLSA**	
		0.6 *threshold*	0.9 *threshold*
Influence / Indépendance *(Impact / Independency)*	496	530	532
Relationnel / Environnement *(Relational / Environment)*	420	468	492
Relationnel / Rôle *(Relational / Role)*	384	**359**	**355**
Rôle / Comportement-Attitude *(Role / Behaviour)*	344	389	**325**
Stress / Indépendance *(Anxiety / Independency)*	481	**392**	**401**
Stress / Vous-même *(Anxiety / Yourself)*	494	**442**	**446**
Vous-même / Comportement-Attitude *(Yourself / Behaviour)*	422	423	**407**

Table 1. LSA and *ExpLSA* with different Asium thresholds (0.6 and 0.9).

Table 1 shows the evaluations obtained on randomly selected concepts for LSA, *ExpLSA* with 0.6 threshold, and *ExpLSA* with 0.9 threshold. We compare the results with a corpus using an Asium threshold of 0.9 versus a large (but less relevant) expansion corpus using a threshold to 0.6. The ranking sums of relevant pairs of terms are compared with LSA. Our *ExpLSA* approach with 0.6 Asium threshold improves the LSA results only 3 times on 7. But when we use a 0.9 threshold, *ExpLSA* improves results 5 times on 7. Thus we achieve better

[2] only the sentences with the instances of concepts are used to compute LSA and *ExpLSA*.

quality results with a 0.9 threshold. However, there are two cases where *ExpLSA* performed badly. They could be studied in a future work.

5 Conclusion and discussion

LSA is a method applied to large corpora. Actually, this analysis is less efficient with small corpora. We study in this paper a corpus to build a conceptual classification. We complete a corpus with our *ExpLSA* approach using syntactic knowledge. Our approach does not improve results for all experiments. However, the results obtained are hopeful. Our experiments have been performed on a small number of concepts. We intend to perform *ExpLSA* with every concepts combination. Moreover, we will estimate more precisely the most appropriate Asium threshold with new experiments.

Acknowledgment: We thank Yves Kodratoff (LRI, France) and Serge Baquedano (PerformanSe company) for their expertise on the Human Resources corpus.

References

1. Landauer, T., Dumais, S.: A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge. Psychological Review, Vol. 104. (1997) 211–240
2. Rehder B., Schreiner M., Wolfe M., Laham D., Landauer T., Kintsch W.: Using Latent Semantic Analysis to assess knowledge: Some technical considerations. Discourse Processes, Vol. 25. (1998) 337–354
3. Wiemer-Hastings P., Zipitria I.: Rules for syntax, vectors for semantics. Proceedings of the Annual Conference of the Cognitive Science Society. (2001)
4. Brill E.: Some Advances in Transformation-Based Part of Speech Tagging. AAAI, Vol. 1. (1994) 722–727
5. Kanejiya D., Kumar A., Prasad S.: Automatic Evaluation of Students' Answers using Syntactically Enhanced LSA. Proceedings of the Workshop on Building Educational Applications using NLP (HLT-NAACL conference). (2003)
6. Bourigault D., Fabre C.: Approche linguistique pour l'analyse syntaxique de corpus. Cahiers de Grammaires, Vol. 25. (2000) 131–151
7. Roche M., Heitz T., O. Matte-Tailliez O., Kodratoff Y.: Exit: Un système itératif pour l'extraction de la terminologie du domaine à partir de corpus spécialisés. Proceedings of JADT'04, Vol. 2. (2004) 946–956
8. Chauché J.: Un outil multidimensionnel de l'analyse du discours. Proceedings of Coling, Standford University, California. (1984) 11–15
9. Faure D.: Conception de méthode d'apprentissage symbolique et automatique pour l'acquisition de cadres de sous-catégorisation de verbes et de connaissances sémantiques à partir de textes : le système ASIUM. Phd, Univ. de Paris 11. (2000)
10. Ferri C., Flach P., Hernandez-Orallo J.: Learning decision trees using the area under the ROC curve. Proceedings of ICML'02. (2002) 139–146
11. Roche M., Kodratoff Y.: Pruning Terminology Extracted from a Specialized Corpus for CV Ontology Acquisition. Proceedings of onToContent'06 workshop - OTM'06, Springer Verlag, LNCS. (2006) 1107-1116
12. Mary J.: Étude de l'Apprentissage Actif : Application à la conduite d'expériences. Phd, Univ. Paris 11. (2005)

Group Profile and Ontology-based Semantic Annotation of Multimedia Data for Efficient Retrieval

Nadeem Iftikhar, Muhammad Abdul Qadir, Omara Abdul Hamid

Mohammad Ali Jinnah University,
Faculty of Engineering and Applied Sciences,
Jinnah Avenue, Islamabad, Pakistan
{nadeem, aqadir, omara}@jinnah.edu.pk

Abstract. Efficient retrieval of multimedia data has gained importance in recent years. There are many techniques for efficient retrieval of textual data; however, not all of them are applicable to multimedia data. The problem of efficiency in the retrieval of multimedia data could be over come by getting the semantics of multimedia data. In this regard, the researchers have adapted the approaches from the domain of image processing and computer vision. Until today, these approaches are not very much matured; therefore, the results which most of the researches wanted could not be achieved. We try to tackle this problem, from our domain of computer science by incorporating group profile and merging the domain and multimedia ontology to annotate the multimedia data semantically. Hence, by semantic annotation, we would able to retrieve the multimedia data efficiently.

1 Introduction

The multimedia data retrieval is very different from textual data retrieval. In textual data, the retrieval is based on keyword/exact matching, whereas in multimedia the retrieval, also known as content-based retrieval (CBR), is mainly dependant on the actual contents of the multimedia data. Various multimedia data items may share the same contents with very minute difference and this difference could only be identified by understanding the semantics of each multimedia data item such as an image. This semantic difference is also called *semantic gap* [1], between the actual multimedia data and human perception about it.

In the past, multimedia retrieval is mainly achieved through low-level features such as color, shape, texture, orientation [2], [3] etc. These features do not provide much help in extracting the semantics of multimedia data. For example, it is very difficult to find *picture of a drawing room* with only low-level features. But, low-level features along with high-level features/annotation can some what achieve it [4], [5] etc. High-level features/annotation can be achieved manually as well as automatically. Manual annotation can provide rich semantics, but it is time consuming and labor extensive. Therefore, it is not feasible to apply it on a large multimedia data set. On the other hand, automatic annotation can overcome these problems, but it may not be able to achieve annotation with affluent semantics.

In our proposed approach, we analyze the multimedia data (image for the time being) though its components. For example if in an image we find *tiger*, *deer* and *tree* or *grassy area* then it means that image is representing the abstract *a tiger is chasing a deer in the jungle*. To tag or name the components we utilize the combination of domain and multimedia ontology, domain ontology for representing high-level features and multimedia ontology for low-level features. And to extract the abstract automatically, we apply the understanding of a specialized group in a community of like minded people.

Use of domain ontology for annotating the multimedia data at the time of storage and later maps the users' query on the same ontology for better results is being proposed by [6] etc. Lux et al. [7] emphasize on applying the standards such as, MPEG-7 for representing the low-level features. Combined approaches of using domain ontology along with multimedia standards/ontology are used by [8] etc. In addition to use a combined ontology Chebotko et al. [9] further added the concept of language profile for making the annotation process personalized and selecting a subset of domain ontology terms for linguistic annotation. The idea of community based profile is being proposed by [10], [11] etc. for sharing and reusing the knowledge within a community of like minded people.

2 Proposed Approach

Our approach is a hybrid approach, which uses the combination of domain and multimedia ontology, almost in the same fashion as used by [8], [9] etc. In addition to the combined ontology, it also uses a group based learning or group profile based approach which is a subset of community based learning used by [10], [11] etc. One of the advantages of group based learning: is the specialized nature of a group as compared with a community of same interests. For example, if we compare the *researcher community* with a specialized research group such as *Database group*, then it is very obvious that the knowledge which we can share or reuse of a specific group will be more precise and accurate to the one from the specific community.

A significant principle behind our proposed approach is to consider the user's context through group profile and annotate the multimedia data automatically by using domain and multimedia ontology along with already stored annotations in the related groups' repository in order to extract the semantics. Our approach consists of six main components: (1) Feature Extractor, (2) Repository, (3) Group profile, (4) Abstract generator (5) Domain ontology, and (6) Multimedia ontology.

Feature extractor extracts the low-level features and fills in the tags of the multimedia part of the ontology. Repository acts like a coordinator between group level stored annotations, abstract generator and group profile. Group profiles contain the information about the group behaviors, restrictions, preferences, history, future events, links to the repository, etc. A number of group profiles are created initially, based on the nature of the users' of the system. On the first use, the user provides his/her profile, which is then analyzed, and the user is associated to a set of related group(s). The group profiles are regularly updated based on the annotations, added by the users of the same group. Group profile will interact with other components at the time of annotation. This is due to the fact that when group profile(s) are included at the time of storage; this can limit the abstract generator to consider only those semantics, which are related to the specific user group at that time. While it is possible that abstract generator finds more than one abstract for the submitted media. If that is possible then image will be annotated with multiple semantics. Group profile along with the annotations already stored in the groups' repository will help towards filling in the tags of the domain part of the ontology. Fig. 1 shows the conceptual level diagram of the approach.

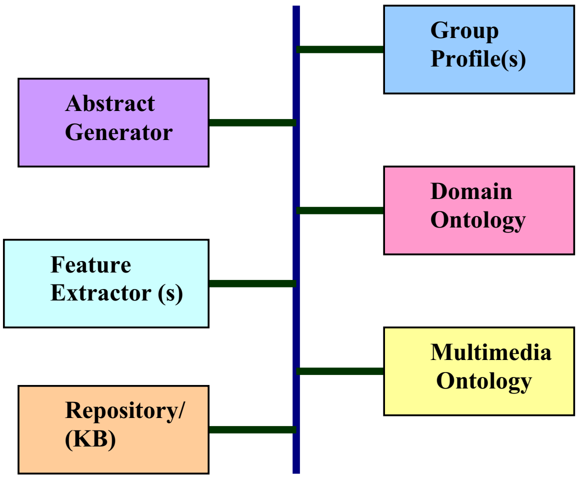


Fig. 1. Main components of the proposed approach

3 Conclusion

Semantic based retrieval of multimedia data depends upon accurate extraction of semantics. It is not feasible to manually write the abstract of ever growing billions of multimedia data available on the *WWW*. Most of the automatic abstract extraction techniques based upon different image processing algorithms use low level features (color, texture, shapes etc), have not come up to the mark yet. We have proposed a novel approach, which is based on combined ontology and group profile. In our approach, automatic abstract generation starts with the extraction of low-level features and then by using group based learning the system narrows down the scope and moves towards the high-level features.

References

1. Dunckley, L.: Multimedia Databases: An Object-Relational Approach. Addison-Wesley Publication, USA (2003)
2. Carson, C., Belongie, S., Greenspan, H., Malik J.: Blobworld: Image Segmentation using Expectation-Maximization and its Application to Image Querying. IEEE Trans. on Pattern Analysis and Machine Intelligence, 24(8), 1026--1038 (2002)
3. Chen, Y., Wang, J. Z.: A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval System. IEEE Trans. on Pattern Analysis and Machine Intelligence. 24(9), 1252--1267 (2002)
4. Kang, F., Jin, R., Chai, J. Y.: Regularizing Translation Models for Better Automatic Image Annotation. In: 13th ACM International Conference on Information and Knowledge Management, pp. 350 --359. ACM Press, Washington (2004)
5. Jeon, J., Lavrenko, V., Manmatha, R.: Automatic Image Annotation and Retrieval using Cross-Media Relevance Models. In: 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 119--126. ACM Press, Canada (2003)
6. Hammiche, S., Benbernou, S., Hacid, M-S., Vakali, A.: Semantic Retrieval of Multimedia Data. In: 2nd ACM Workshop on Multimedia Databases, pp. 36--44, ACM Press, USA (2004)
7. Lux, M., Klieber, W., Becker, J., Tochtermann, K., Mayer, H., Neuschmied H., Haas, W.: XML and MPEG-7 for Interactive Annotation and Retrieval using Semantic Meta-data. Journal of Universal Computer Science. 8(10), 965--984 (2002)
8. Petridis, K., Bloehdorn, S., Saathoff, C., Simou, N., Dasiopoulou, S., Tzouvaras, V., Handschuh, S., Avrithis, Y., Kompatsiaris I., and Staab, S.: Knowledge Representation and Semantic Annotation of Multimedia Content. Vision Image and Signal Processing, Special issue on Knowledge-Based Digital Media Processing. 153(3), 25--262 (2006)
9. Chebotko, A., Deng, Yu., Lu, S., Fotouhi, F., Aristar, A.: An Ontology-Based Multimedia Annotator for the Semantic Web of Language Engineering. Int'l Journal on Semantic Web & Info. Sys. 1(1), 50--67 (2005)
10. Chakravarthy, A., Lanfranchi, V., Ciravegna, F.: Community-Based Annotation of Multimedia Documents. In: 3rd European Semantic Web Conference. Montenegro (2006)
11. Kuflik, T., Shapira, B., Shoval, B.: Stereotype-Based versus Personal-Based Filtering Rules in Information Filtering Systems. Journal of the American Society for Information Science and Technology. 54(3), 243--250 (2003)

Between symbol and language-in-use

Emma Tonkin

UKOLN,
University of Bath, BA27AY, UK
e.tonkin@ukoln.ac.uk
http://www.ukoln.ac.uk/

Abstract. Indexing is often designed with the intent of dimensional reduction, that is, of generating standardised and uniform descriptive metadata. This could be characterised as a process of decontextualisation. Formal knowledge representation systems typically have the aim of encapsulating granular pieces of information in a reusable manner. The result is a set of information elements with minimal links to external information sources. Plain-text tags, by comparison, have the aim of describing an object, within or outside a reductively described context. The result is a set of views that are contextualised to author, time, location, task or community. This paper discusses the relationship between symbol, contextual relation and language-in-use.

Key words: Language-in-use, metadata analysis, informal metadata

1 Introduction

Formal knowledge representation systems are often reductionist in philosophy; that is to say, a goal of techniques such as metadata collection is often the establishment of a dimensionally reduced set of data records from which to operate. In this sense, we might cast the process as one inherently concerned with decontextualisation, occurring in the Aristotelian tradition[5]. All extraneous variables are normalised to default values. Whether a domain specific encoding is used or a generic sub-language, the derived result is enforced homogeneity. The aim, though it is contentious to what extent the aim is practically realised, is generally to relate objects (physical, electronic or conceptual) to a formally defined model.

This paper is intended to contribute to an existing discussion regarding the process of developing a formal system and the gradient between natural language, metadata and formal system. This is of relevance to a number of topics in information retrieval. Formalisation in knowledge representation and retrieval is a mature topic. Wilks (2006) discusses the relationship between natural language and formalism in terms of the Semantic Web, describing 'two differing lines of Semantic Web research: one, closely allied to notions of documents and natural language (NL) and one not.' Here, we ask a similar question in terms of semi-formal and informal metadata in use.

1.1 Representation in natural language

One might begin by asking the simplest of questions: What is a language? As a functional definition, a language allows one to speak and be understood by others who know that language [3]. It is possible to describe a language as a system composed of symbols with referential meaning; however, such a description does little to clarify that which is intended. Deacon [2] articulates two definitions of *symbol*, one drawn from the humanities and another drawn from computation.
Humanities: A symbol is one of a conventional set of tokens that marks a node in a complex web of interdependent referential relationships and whose specific reference is not obviously discernible from its token features. Its reference is often obscure, abstract, multifaceted and cryptic, and tends to require considerable experience or training to interpret.
Computation: A symbol is one of a conventional set of tokens manipulated with respect to certain of its physical characteristics by a set of substitution, elimination and combination rules, and which is arbitrarily correlated with some referent.

To precise further, then, would place this paper on one side or another of an interdisciplinary rift - a dichotomy both apparent and keenly felt. What is the validity of this voluntary separation between disciplines?

Deacon characterises these uses as 'complementary, referring to two different aspects of the same phenomenon', with the computational definition predominantly describing the production and manipulation of the symbol tokens and the former definition relating rather to the interpretation and symbolic effects of symbolic reference. The distinction is drawn between icon and system – an icon could be said to form a reference by drawing on similarity to a referent, a resemblance, whilst a symbolic token has no such constraints, though it may act as an icon in certain contexts. However, the relevance of iconicity in acquisition of a symbol is questionable (see [8], p.36-37). Links between symbol and world are drawn by indexical reference - a mapping between symbol and referent, based on correlation.

1.2 Natural language to formal representation

There exist formal treatments of natural languages, characterising NL as a set of algebraic rules and a lexicon of meaningful linguistic elements[6]. There exist also usage-based theories of linguistics that treat structure as resultant (emergent) from language use [8].

One could place various approaches toward information representation on a spectrum of increasing regularity or completeness of intended definition of symbol and/or relation. Alternatively, one could order systems according to the set of assumptions which underlie each approach; for example, some simple statistical systems rely simply on term frequency/keyword density, with others depending instead on the distributional hypothesis. Metadata schemas are often designed explicitly for use within a single environment, with a more or less completely defined set of use cases. Context is explicitly handled by metadata application profiles, which provide a means of labelling records as resulting from a given application type of a metadata schema.

In practice, a representation designed for information management and retrieval purposes is typically influenced by concerns other than cognitive or neural realism. Computability, for example, is a primary concern. An optimal representation may therefore be far from realistic. From the HCI viewpoint, a probable approach for creating such a representation is likely to make use of information elicited from study of appropriate stakeholders. An alternative approach is the encoding and use of an existing formal representation. Either way, a formal classification is far from decontextualised; to quote Stephen Jay Gould on the purpose of classification: "[Classifications represent] theories about the basis of natural order, not dull categories compiled only to avoid chaos."

The development of a formal representation involves decontextualisation in the sense that extraneous contextual information is given only implicitly. The context of a formal ontology is formally given only in the sense that a namespace is provided; the syntax rules and lexicon are provided with the implicit datum that they apply only within the operative context of this representation. However, this contextual information is not explicitly encoded – which statement is not intended to suggest that it is possible or desirable to do otherwise. The purpose of the process of formalisation is generally reductionist. The eventual aim is the extraction of information in a form usable for the diverse purposes of the system, which implies the need to collect information in a form appropriate for that purpose.

Deacon[2] notes Frege's recognition that 'words on their own generally do not refer to particular concrete things in this world except when in certain combinations or contexts that determine this link'. Brief utterances require explicit context to be appropriately interpreted. What explicitly given context has a key-value pair in, for example, Dublin Core metadata?

Applied language – language-in-use – acquires a syntax and semantics characteristic of its domain of use and of the actors between which the term is used. The design of formal systems for each knowledge subdomain or scenario represents a formal (analytical) approach to the same representation task that language systems in general approach in a more general manner. This leads us to a question that might be described as a recurring theme in digital library research: why, in a given scenario for metadata use, would we expect a formal system to be more "appropriate" than a system developed by participants in the process for use in the area – and by what metrics might we measure appropriateness?

2 Metadata as language-in-use

An approach toward formal representation of information is no more accurate than the users who apply it. Wilks[9] points out that this is a standard philosophical problem; as annotations are used to bind text to meaning representations, the markers themselves are said by some critics to take up the characteristics of natural language and therefore reach no meaning outsid language. As a solution, linking the virtual world to real-world quantities and artifacts is suggested. Though accepted as a plausible approach, Wilks adds that 'Nothing will satisfy a critic... except a web based on a firm (ie. formal and extra-symbolic) semantics and effectively unrelated to language at all [...] The SW may be the best way of showing that a non-formal semantics can work effectively, just as language itself does and in the same way.'

This is the most revealing of quotes. If designers of informal and semi-formal semantic systems are building languages, then that language may be expected to be as susceptible to contextualisation in speech acts as any other. Is it possible that appropriate analysis of real-world use of existing indexing systems (in the sense of annotation systems rather than textual analysis approaches - although application of such techniques may well qualify as 'appropriate analysis') would show that contextualised use of metadata is already with is, though the encoding is not an explicit one?

2.1 Tagging systems

Plain-text tags have the aim of describing an object and providing a pointer to that object - the generation and use of free-text metadata for description and discovery of resources. The result is a set of views that are contextualised to author, time, location, task or community. The tag is the vaguest of indexing systems. A tag corpus is constructed of a set of speech acts, and each term is generally devoid of context in the sense of grammar or syntactic relatives. The relative of the distributional hypothesis in tagging could better be labelled the " co-occurrence hypothesis" - similar words are preferentially used to point to similar items.

Tags are simply snatches of natural language, though some efforts have been made to encourage consistent use of conventions such as spelling, pluralisation and so forth. There is an argument to be made that tags are simply keywords, and indeed the difference is more likely to be found in the domain of use and characteristics of the user community than in the technology itself. Either way, tag corpuses provide a fascinating opportunity to examine a largely user-driven adaption of natural language for indexing purposes. Any reductionist influence present in this subset of language exists either due to technical limitations or the decision of the individual providing the tag. This provides for the fascinating possibility that a limited subsystem of language can arise from applied use of natural language in a given context – a folksonomy. The characteristics of such a sublanguage are in general studied, rather than as a corpus of interest to linguistics, as a keyword corpus in need of filtering.

2.2 Evidence from semi-formal metadata

It is undoubtedly easy to point to patterns of failure in the application of semi-formal metadata systems, such as for example Dublin Core application profiles in a given information retrieval context. However, to pinpoint the causes of such failures is relatively difficult. It is probable that they frequently result from problems such as ambiguity in key names as interpreted by the user community – that is, misunderstanding of the intended use of a given field – and changes in the scope of use of a given schema following its introduction.

One might describe this, somewhat flippantly, as analogous to the Whorfian hypothesis in action. Where the hypothesis suggests that one cannot think something for which one does not possess a linguistic representation, this relates instead to the assumption that one cannot represent something for which one does not have an appropriate element in one's schema. This, of course, is false; in practice user populations typically manage very well in the face of unexpected requirements, sacrificing interoperability by applying a sensible-sounding self-sponsored adaption to the system. Such adaptions may be characterised according to many factors. Drivers such as intended audience and convenience are significant. What prompts such inventions, and under what circumstances does the motivation for incorporating an original concept overcome deterrent factors?

Tennis[7] notes that tagging "seems intensely personal, whereas subject cataloguing is an act of delegation mediated by institutions," drawing a clear distinction between indexing as a prescriptive and as a descriptive process. Both processes take place in a definable context - in the first, the context is personal ; in the second, it is institutional. The intended audience of descriptive speech has a significant impact on the ease by which it may be interpreted after the fact (see for example the experiment described by Lave[4]). With this and the earlier discussion of classification as theory in mind, it seems appropriate to ask whether the construction of many current information retrieval systems does not already amount to a set of suppositions regarding the context of use.

To come to an understanding of the domains in which an information retrieval system succeeds or fails is a special case of a general problem; that of the appropriateness of a symbolic system for a given case. The handling of context in natural language itself is far from simple, though it may be modelled in a number of ways, such as by application of variants on the distributional hypothesis. Language in general carries various indicators of context on syntactic and semantic levels. It is reasonable to expect that in practical application, formal symbolic systems will acquire (and very probably already exhibit) a very similar character.

3 Conclusion

The possibility of formally encoding the notion of context inspires a counterpoint question – is it possible to formally exclude the notion? The capacity for creating a formal representation does not necessarily imply that such a representation is wanted or needed; there is an argument to be made that arbitrarily created representations are theories, ways of classifying the world around us. In many cases, it is likely that the need for the structures themselves is not as yet ascertained.

The design of information retrieval systems is complicated by a number of factors, one of which is the difficulty of establishing situationally appropriate metrics for evaluation. Ultimately, the question to be asked may be *why?* Natural languages can perhaps be characterised as compromising between a variety of competing aims, and artifically created or defined languages may be characterised similarly.

A current aim of our research is to examine existing corpora of informal and semi-formal metadata and, from this information, to characterise present patterns of use of these approaches. We find it probable that for our purposes, the simplest approach to contextualised metadata is to work as far as possible with the markers already present in indexing data. To examine the process of creation and use of an existing corpus of data may tell us more about what is already encoded or may be retrieved from the dataset – at the least, this approach may prove beneficial from a vocabulary management perspective.

References

1. Bearman, D., Trant, J. (2005), Social Terminology Enhancement through Vernacular Engagement Exploring Collaborative Annotation to Encourage Interaction with Museum Collections D-Lib. **11** 9, doi:10.1045/september2005-bearman
2. Deacon, T. W. (2003) UG and Semiotic Constraints. In: Language Evolution. Ed: Morten H. Christiansen and Simon Kirby. Oxford.
3. Fromkin, V., Rodman, R. (1993) An Introduction to Language, Fifth Edition. Harcourt Brace Jovanovich.
4. Lave, J. (1991), *Situating learning in communities of practice.* In: L.B. Resnick, J. Levine and S. D. Teasley (Eds.) Perspectives on socially shared cognition. Washington, DC: American Psychological Association.
5. Peterson, E. (2006), Beneath the Metadata: Some philosophical problems with folksonomy. D-Lib. **12** 11, doi:10.1045/november2006-peterson
6. Pinker, S. (1999), *Words and Rules: The Incredients of Language.* New York: HarperCollins.

7. Tennis, J.T. (2006) Social Tagging and the Next Steps for Indexing. In Furner, Jonathan and Tennis, Joseph T, Eds. Proceedings 17th SIG/CR Classification Research Workshop, Austin, Texas.
8. Tomasello, M. (2003), Constructing a language: A usage-based theory of language acquisition. Harvard University Press.
9. Wilks, Y. (2006) The Semantic Web as the apotheosis of annotation, but what are its semantics? International Journal of Web Semantics

RECENT RESEARCH REPORTS

#116 Marco Baroni, Alessandro Lenci, and Magnus Sahlgren, editors. *Proceedings of the 2007 Workshop on Contextual Information in Semantic Space Models: Beyond Words and Documents*, Roskilde, Denmark, August 2007.

#115 Paolo Bouquet, Jérôme Euzenat, Chiara Ghidini, Deborah L. McGuinness, Valeria de Paiva, Luciano Serafini, Pavel Shvaiko, and Holger Wache, editors. *Proceedings of the 2007 workshop on Contexts and Ontologies Representation and Reasoning (C&O:RR-2007)*, Roskilde, Denmark, August 2007.

#114 Bich-Liên Doan, Joemon Jose, and Massimo Melucci, editors. *Proceedings of the 2nd International Workshop on Context-Based Information Retrieval*, Roskilde, Denmark, August 2007.

#113 Henning Christiansen and Jørgen Villadsen, editors. *Proceedings of the 4th International Workshop on Constraints and Language Processing (CSLP 2007)*, Roskilde, Denmark, August 2007.

#112 Anders Kofod-Petersen, Jörg Cassens, David B. Leake, and Stefan Schulz, editors. *Proceedings of the 4th International Workshop on Modeling and Reasoning in Context (MRC 2007) with Special Session on the Role of Contextualization in Human Tasks (CHUT)*, Roskilde, Denmark, August 2007.

#111 Ioannis Hatzilygeroudis, Alvaro Ortigosa, and Maria D. Rodriguez-Moreno, editors. *Proceedings of the 2007 workshop on REpresentation models and Techniques for Improving e-Learning: Bringing Context into the Web-based Education (ReTIeL'07)*, Roskilde, Denmark, August 2007.

#110 Markus Rohde. *Integrated Organization and Technology Development (OTD) and the Impact of Socio-Cultural Concepts — A CSCW Perspective*. PhD thesis, Roskilde University, Roskilde, Denmark, 2007.

#109 Keld Helsgaun. An effective implementation of k-opt moves for the Lin-Kernighan TSP heuristic. 2006, Roskilde University, Roskilde, Denmark.

#108 Pernille Bjørn. *Virtual Project Teams — Distant Collaborative Practice and Groupware Adaptation*. PhD thesis, Roskilde University, Roskilde, Denmark, 2006.

#107 Henrik Bulskov Styltsvig. *Ontology-based Information Retrieval*. PhD thesis, Roskilde University, Roskilde, Denmark, 2006.

#106 Rasmus Knappe. *Measures of Semantic Similarity and Relatedness for Use in Ontology-based information Retrieval*. PhD thesis, Roskilde University, Roskilde, Denmark, 2006.

#105 Davide Martinenghi. *Advanced Techniques for Efficient Data Integrity Checking*. PhD thesis, Roskilde University, Roskilde, Denmark, 2005.