# 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.

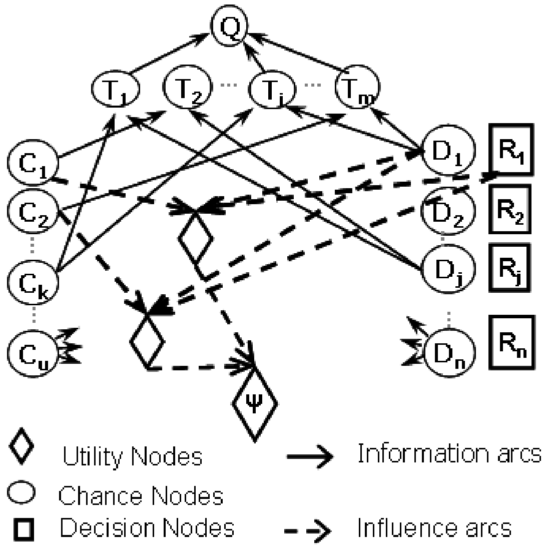

**Fig. 1.** Influence diagram-based retrieval model

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 $\theta$), we can compute the probability function attached to a query node using the *fusy-Or* aggregation operator [10] such as:

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

  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:

$$P(t_i/d_j, c_k) = P(t_i/d_j) * P(t_i/c_k) \qquad (2)$$

  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:

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

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

  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 $C_k$, $\delta$ and $\gamma$ constant values $(0 \le \delta, \gamma \le 1)$.
  More precisely:

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

  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 $\theta$ represents the whole possible configurations of the terms in $pa(Q)$, $\theta^s$ the $s$ order configuration, and $\theta_i^s$ 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

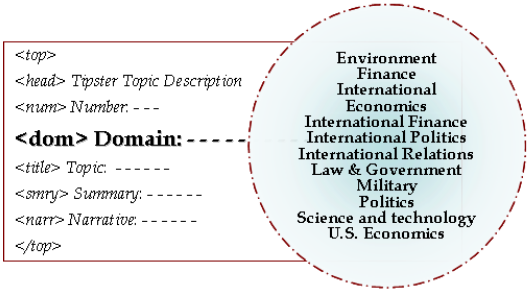

**Fig. 2.** A TREC query annotated with domain meta data

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