Towards an architecture for personalized information retrieval implying the user's profile and votes

Azza Harbaoui, Sahbi Sidhom, Malek Ghenima, Ghezala Henda Ben

Towards an architecture for personalized information retrieval implying the user's profile and votes

Harbaoui Azza#1, Sahbi Sidhom#2, Malek Ghenima#3, Henda Ben Ghezala#4

1 #3 #4 *RIADI Laboratory, ENSI, University of La Manouba, La Manouba 2010, Tunisia*

#2LORIA Laboratory & Nancy University,Vandoeuvere 54506,France

1azza.harbaoui@gmail.com, 2Sahbi.Sidhom@loria.fr, 3malekghenima@gmail.com, 4Henda.BG@cck.rnu.tn

Abstract— Information Retrieval is a process made by a user to obtain relevant information which meets his needs using an Information Retrieval System (IRS). However the IRS shows some differences between user relevance and system relevance. These variations are primarily related to the imperfection of the indexing process (approach directly related to the IR) and specially the non consideration of the user profile. This paper presents a study about formalisms proposed in the literature and addresses some reflections to deal with problems arising from this survey, in order to satisfy the final user in Information Retrieval process.

Keywords— information retrieval, user profile, personalized IR, indexing, user re-indexing, ontology

I. INTRODUCTION

As an Academic field of study Information Retrieval (IR) might be defined thus *"Information Retrieval is finding material (usually documents) of an unstructured nature (usually text) that satisfies an information need from within a large collections (usually stored on computer)"* [2]. Furthermore, in the case of textual documents in which we are interest a significant part of difficulties are due to the difficulty lies in ambiguity inherent to human languages [1] and the carelessness of the user. As a consequence, we observe that in order to build more robust Information Retrieval System (IRS) able to interact naturally with human, we should imply the user. The work presented in this paper deals with an overview of IR. In section III, we present the main information retrieval models in section IV we present the evolution of the classic information retrieval to adaptive information retrieval, our work in progress, proposing some key ideas in order to improve information retrieval. Finally, we conclude this paper.

II. FOUNDATIONS OF CLASSIC INFORMATION RETRIEVAL

In this section, we present the foundations of classic Information Retrieval and an outline of the main models.

According to Rijsbergen who defines Information Retrieval as *"The user expresses his information need in the form of a request for information, Information Retrieval is concerned with retrieving those documents that are likely to be relevant to his information need as expressed by his request"*. This definition contains two important things that we should define: the document and the request.

- The document: we call document any unit of information which can constitute an answer to a user's request. It can be a text, a part of text, a picture, a video band
- The Request: constitutes the expression of the need in user's information, the user has to subject to the search engine his need in information. Diverse types of query languages have been proposed in the literature:
  - List of keywords: SMART system and OKAPI [3], [10]
  - Natural language: SMART system and SPIRIT [4]
  - Boolean language: DIALOG system [5]
  - Graphic language: NEURODOC [4]

The principle goal of an information retrieval system is to select the nearest documents that answer a user request. For that purpose, the information retrieval system regroups a set of methods and procedures allowing the management of the collections of documents stored in the form of an allowing intermediate representation.

Thus, the interrogation of the collection of documents by means of request requires the representation of this last one under a shape one unified compatible with those of documents. These features are represented with a global process of IR, collectively named process in U illustrated in figure 1.

This process consists in two main phases: the indexing and the interrogation.

***A.* The indexing process:** it's a very important step in the process of information retrieval it consists of determining and extracting the representative terms of a document or request. The result of the indexing process is called descriptor: it can be a list of terms, or a set of significant terms. The descriptors are grouped in a catalogue called "Dictionary». This dictionary constitutes the language of indexing process.

There are different types of indexing: manual indexing, semi-automatic indexing and automatic indexing.

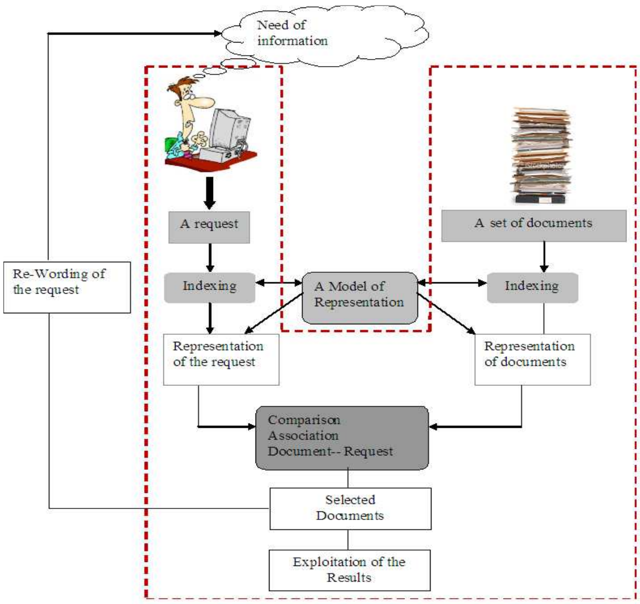

Figure 1: Information Retrieval process (U form)

- Manuel or intellectual indexing: an expert chooses terms that he judge relevant in the description of the document content. Furthermore, this approach is subjective on the one hand because it depends on the expert's knowledge, and not very efficient for voluminous collection.
- Automatic indexing: It's based on algorithms that associated automatically descriptors for parts of documents.
- Semi-automatic indexing: it's a combination of the two previous processes: a first automatic process for extracting terms form documents, then a specialist of the domain will use a vocabulary of thesaurus or a terminological base.

According to [2], the basic techniques of automatic indexing are made according to these steps:

1. Collect the documents to be indexed
2. Tokenize the text
3. Do linguistic pre-processing of collections
4. Index the documents that each term occurs in

In order to find the terms that represent best the semantic contents [2] has defined term frequency (TF) and Inverse Document Frequency (IDF).

- TF: This measure is proportional with the frequency of the term in the document: the idea is that more the term is frequent in the document, more it's important in the description of the document.
- Inverse Document Frequency: it measures the importance of the term in all the collection: the idea is that the terms that less appears in the collection are more representative of the document content than those that appear in all the documents collection.

B. The Interrogation mechanism:

The interrogation system implies a process of interaction with the user illustrated in figure 1.In fact this interaction includes the formulation of the request by the user that translate his need in information, the representation of the query in an internal form according to a defined indexing language, the correspondence between the request and the documents. More precisely, the interrogation implies the following scenario:

The user expresses his need of information with a request. The system will index the request and create a request index which is compatible with the documents index. Then the system evaluates the relevance of the documents comparing to the request using a function of correspondence. This function interprets the generated index in order to calculate a score of similarity called "Relevance Status Value" (RSV).

Thus, several kinds of models in information retrieval have been proposed in the literature trying to formalize the relevance starting from basic models towards more elaborated ones.

III. INFORMATION RETRIEVAL MODELS

According to [6], a model gives a formalization of the information retrieval process as well as a theory for the modelling of the measure of relevance.

Because of space limitations, we will not present all the models. References are provided to give further information concerning these works.

– Boolean Model

This model was proposed to represent the request in the form of logical expression. The indexing terms are connected with logical connectors (AND, OR, NOT). A Boolean model only records term presence or absence in order to realize an exact matching with the equation of the request. The documents that satisfy the logical request's expressions are considered as relevant.

Thus the Retrieval Status Value (RSV) is the logical result of $d \rightarrow q$ (d: document, q: request) in the extension of this model, the selection of documents is based on approximately matching not exact one. The Boolean model is easier to implant and requires relatively few resources. Furthermore, it is sometimes difficult for the user to express his information need's with boolean expression in particular for the complex request, what does not allow to make best use of the model's features.

– Vector Space Model

Inspired by Luhn [8], the vector space model, introduced by Salton [4] in the SMART project (Salton's Magical Automatic Retriever of Text).This model is based on the mathematical bases of vectorial spaces. The request and the document are represented in the vector space generated by indexing terms. In fact, the degree of relevance of a document toward a request is proportional to the position of these vectors in the space. It's evaluated with a correlation degree correlation between associated vectors. The similarity measure between documents and requests can be made of several manners: for example by a scalar product, by a measure of train which forms vectors, by the measure of the angle which forms the vectors, or still by a measure of distance.

The measure of similarity SIM(D, Q) can be calculated according to:

$$SIM(D_j,Q)=\sum_{i=1}^{n} d_{ij}.q_i \quad \text{(Scalar Product)}$$

Or

$$SIM(D_j,Q)=\frac{\sum_{i=1}^{n} d_{ji}\bullet q_i}{\sqrt{\sum_{i=1}^{n} d_{ji}^2 \bullet \sum_{i=1}^{n} q_i^2}} \quad \text{(Cosine)}$$

We can also use a measure of distance:

$$DIST(D_j,Q)=\sqrt{\sum_{i=1}^{n}(d_{ji}-q_i)^2}$$

(Euclidian distance).

Other measures of similarity are useful, being situated halfway enter the vector space model and the Boolean model such as:

$$\text{- } DiceCoefficient:(D_j,Q)=\frac{2\sum_{i=1}^{i=n} d_{ji}\times q_i}{\sqrt{\sum_{i=1}^{i=n} q_i}+\sqrt{\sum_{i=1}^{i=n} d_{ji}}}$$

- Jaccard Coefficient

$$JACCARD(D_j,Q)=\frac{\sum_{i=1}^{n} q_i.d_{ij}}{\sum_{i=1}^{n} q_i^2+\sum_{i=1}^{n} d_{ij}^{\;2}-\sum_{i=1}^{n} q_i.d_{ij}}$$

- Pseudo cosine

$$pseudo-\cos(D_j,Q)=\frac{\sum_{i=1}^{i=n} d_{ij}\times q_i}{\sqrt{\sum_{i=1}^{i=n} q_i}+\sqrt{\sum_{i=1}^{i=n} d_{ij}}}$$

-Measure of recovery: $recouvr(D_j,Q)=\dfrac{\sum_{i=1}^{i=n} d_{ji}.q_i}{\min(|D_j|,|Q|)}$

The advantages of the vector space model are many:

- Conceptual Simplicity and implementation
- Uniform representation of the document and query
- Results are ranked in order of relevance using a similarity measure.

Many methods of sequencing results were compared with the vector model, and the later, despite its simplicity, is higher than or at least as good as the other alternatives. For all these reasons today the vector model is most popular in information retrieval. Furthermore, the vector space model has the "theoretical" inconvenient that all the terms of the index are independent, in practice, the global consideration of the terms dependence can make lower the quality of the system's answer (because dependence are in general local).

– *Probabilistic Model*

This model [14], [15] treats the problem of information retrieval in a probabilistic way. The relevance document-request is translated by the calculation of the probability of relevance of the document with regard to a request.

The probabilistic approach was developed by *Turle* et *Croft* in INQUERY, where the indexing uses the probabilistic model. In order to calculate the probability that the relevant event arrives knowing the formulated request *(Q)* and the Document (D), INQUERY uses Bayesian network by taking into account dependences between certain events. It's advantage that it takes into account in particular between the document and its representation.

Probabilistic model are focused on document and query. It uses simple framework and efficient computation time and performance. Furthermore, distribution of the terms for relevant and irrelevant documents should be known or requires labelling.

– *Language Model*

It's a statical language model assigns a probability to sequence of *m* words $P(w_1, \ldots, w_m)$ by means of a probability distribution. This model tries to capture the proprieties of a language and to predict the next word in a speech sequence. When used in information retrieval, a language model is associated with a document in a collection. With a query *Q* as input, retrieved documents are ranked based on the probability that the document's language model would generate the terms of the query, $P(Q|M_d)$.

– *N-gram model*

It's a type of probabilistic for predicting the next item in such a sequence. N-gram models are used in various areas of statical natural language processing (NLP) and genetic sequence analysis. In an n-gram model, the probability $P(w_1, \ldots, w_m)$ of observing the sentence $w_1,\ldots,wm$ is approximated as:

$$P(w_1,\ldots,w_m)=\prod_{i=1}^{m} P(w_i|w_1,\ldots,w_{i-1})\approx\prod_{i=1}^{m} P(w_i|w_{i-(n-1)},\ldots,w_{i-1})$$

The language models provide an interesting method for developing information retrieval in a well-defined theoretical framework. Although, language models require a Part-Of-Speech-Tagging by assigning each word with its grammatical category. As a language model is a probabilistic concept, it will determine the probability of each bigram that appears in a given text.

In fact, the common point to all the models in classic Information Retrieval is that all the selected documents has to contain all the words(or a part of), formulated by the closest semantically to the user's need. Thus, the relevance of the selection relies mainly on the quality of the indexing and matching mechanism. Nevertheless, in the practise the majority of the requests expressed by the users are ambiguous [11] and short, what gives incoherent specification onto their need in information

Besides, the list of terms does not correspond necessarily to those used to index the relevant documents of the collection and always miss a significant term that really expresses the user's need. This leads to crucial problems in classic information retrieval: the first problem is called *"term mismatch problem"*.

The second problem is ambiguity [7]: the user and the author of the document do not use the same vocabulary, so that a document can be relevant even if it does not contain the same terms as those of the request. However, the classic information retrieval system answers invariably the users by sending back the same results to two different users having expressed the same request but having different needs in information and different preferences.

Therefore, the performance of an information retrieval system, do not depend only on the efficiency and quality of mechanisms of indexing and matching but also on the capacities of interaction with the user in order to better satisfy what his expectations.

Of this report appears the adaptive information retrieval.

IV. FROM CLASSIC INFORMATION RETRIEVAL TO ADAPTIVE INFORMATION RETRIEVAL

The goal of the adaptive information retrieval is to adapt the process of information retrieval in order to return relevant results to the user. Techniques developed in adaptive information retrieval [18] focus mainly on the assistance in the reformulation of the request.

1. *Reformulation of the requests*

It aims to generate a new request without explicit retroaction of the user, just by exploring the first documents presented by the system as an answer to the user's request (blind feedback).Different techniques were introduced in the literature, in models of information retrieval such as Vector Space Model [12], and the probabilistic model. We quote the techniques of the reformulation of the requests by relevance feedback, techniques of disambiguation of the request's sense or the techniques of clustering.

2. *Disambiguation of the request's sense*

The aim of these techniques is to assist the user in order to better express his need by adding additional semantic resources [19] (dictionaries, networks, thesauruses).Most of them are based on exploring interactive interfaces of clarification based on ontology

The figure 2 shows an example of an ambiguous query (apple) on Google and the semantic categories that offers Google to the user in order to assist him to find the expected result. In this example, the user is searching for apple which is an ambiguous request because "apple" can be "Apple iPod", "apple fruits", Google proposed disambiguation of the request and the user choose apple fruits from the list proposed by Google

3. *Techniques of Clustering*

In order to face major growth of the web, and the difficulties faced by the classic search engines, the aim of the clustering techniques [20] is to group the results into categories such as Grouper [13], Vivissimo, Kartoo, as well as techniques of repertorisation of the web into taxonomy of concepts (Google hierarchy).All these techniques were developed in order to satisfy the user and to simplify his navigation.

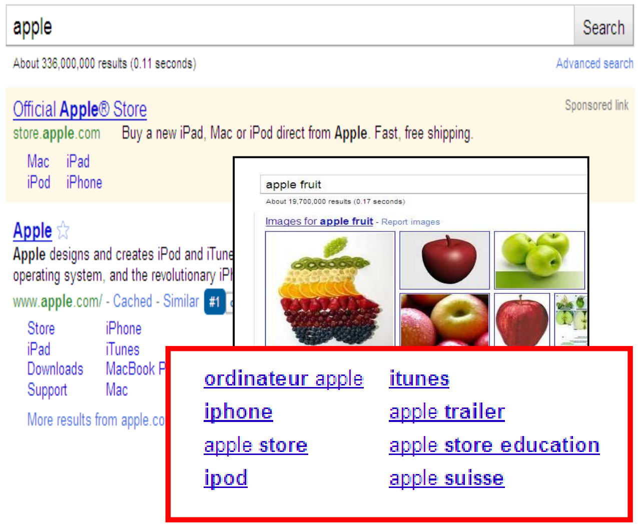

Figure2: An example of request's disambiguation

Thus, the users will have a virtual classification of the documents. Several ontology of domain has been conceived in order to build Conceptual information retrieval.

The theoretical study that we carried out reveals that beside the earnings bring by the techniques of adaptive information retrieval, there still some limitations that we explicit in these points:

i) The context: The context in adaptive information retrieval is not well defined the techniques of reformulation of the requests have to help the user to select documents via interfaces that clarifies the information needs. The context is limited with the terms that the user uses to express his need in information Thus the context in which the research is made is rarely used to interpret the request (the geographic situation for example) .

ii) Limitations due to the explicit interaction: The techniques of reformulation of the request by relevance feedback need an explicit interaction with the final user. Besides, among the studied works, the majority of the users are not motivated to interact with the system, they prefer the automatic mechanism that ameliorates the results without asking for additional information

iii) The impact of the familiarity of the user with the subject:

The relevance of the techniques of the query's reformulation depends relatively on the one hand on the level of familiarity of the user with information retrieval system [22] and on the other hand on the number of iterations of research.

Although the limitations of the adaptive information retrieval, it has certainly brought solutions in particular in the mechanism of matching document and request and always ameliorates the performances of the information retrieval process. However, fine analyses of the domain shows the adaptive information retrieval depends on many factors described above.

Thus, in order to ameliorate the performances of information retrieval system, many studies focused on the impact of the cognitive dimension of the context defined with the user's interest and his preferences.

V. CONTEXTUAL AND PERSONALIZED INFORMATION RETRIEVAL

According to [16], the contextual information retrieval is defined as *"combine search technologies and knowledge about the query and user context into a single framework in order to provide the most appropriate answer for a user's information needs"*.

In the literature [9], the context of research is bound to dimensions dependent to the user, the request and the environment of research. We are interested in personalized information retrieval which is a specific branch of contextual information retrieval. The goal of personalization in information retrieval is to tailor the search engine results. The first approaches of the contextual information retrieval have focused on the context of the user represented by his profile.

This context includes his interests, his votes, his comments, and also his knowledge. In fact people's future preferences and needs must be able to personalize services, the same way recommender systems do. The personalization is a process which changes the feature, the interface, the content information or the aspect of a system in order to increase his personal relevance related with the socio-demographic characteristics declared by the user and his observed behaviour contained in what we call user model.

This model describes any information about the user, as his preferences, his need in information and also his environment of research. The applications are diverse: recommender systems, filtering systems, learning systems as well as information retrieval systems. Independently of the aimed application objective, we identify three main aspects to be promoted in the personalized access systems:

- The capacity to identify the abstract intention of the user
- The flexibility of the process selection in order to adapt itself to the common context of the user
- The interactions between the user and the system

Thus, the relevance of these systems depends on the precision of the exploited user models. Indeed the modelling of the profile constitutes the essential element in the development of a successful personalized system.

To reach this goal, we are now focusing on the construction of the user profile, and its representation in order to ameliorate the information retrieval. The profile [17] can reflect the interest of the users in several subjects at one particular moment. It's a database where the user information, interests and preferences are stored.

A general improvement of personalization systems would be taking into account more semantics in the process. This aspect has to be considered in the construction of the profile (and thus in the user model).

One way to deal with this task is to consider some results coming from the "Indexing Process" and try to ameliorate it by implying the user (his comments, his history, his virtual communities...). In fact, the idea is to integrate the user profile in the indexing process in order to achieve a process that we called "Re-indexing".

We implement the first part of our architecture which is the baseline system as shown in Figure 3.

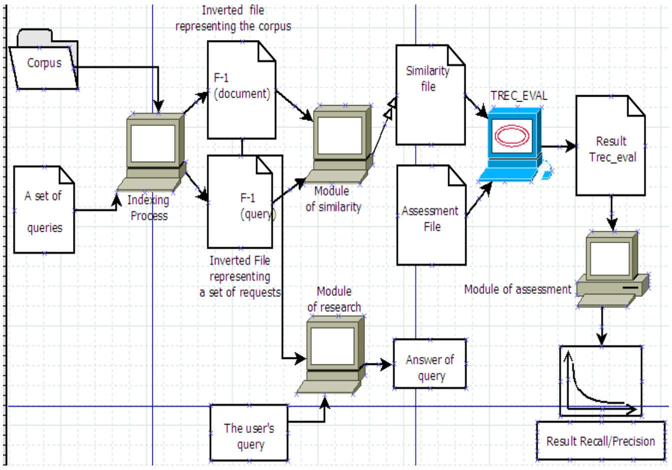

Figure 3: Architecture of the Baseline System

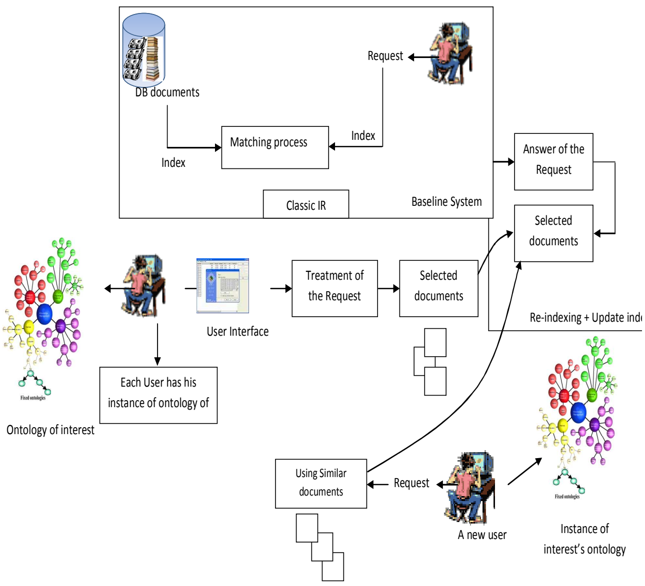

Figure 4: Proposed Architecture of our system

VI. WORK IN PROGRESS

The reflections presented in this paper are situated precisely in the context of personalized access to information. We are now focusing on the construction of architecture, as shown in Figure 4.

The objective is to add semantic value to information retrieval systems toward a new indexing process or re-indexing of the contents. We are interested to imply the user, his profile, his traces, his culture, and his point of view in order to collect all these points into ontology of interest.

In the proposed architecture we start from a classic information retrieval: we want to proceed first with the Baseline System (Figure 4) that we have already implement observing the results in order to compare after with the process of “re-indexing”. We focus on the indexing process by adding semantic values in order to re-index the contents by the votes of the users.

In fact, first a user is connected via a user interface, asking for information with a query, the information retrieval system will treats his need and send to the user the documents selected. Ontology of interest will be created and updated later.

When a new user is connected to the system, his ontology of interest is created too, and the system will recommend to him previous selected documents that matches the best his profile.

Our architecture is divided into two main parts: The classic information retrieval (Baseline System) and the personalized IR(which includes the Re-indexing process) where the user is involved with his profile.

In the first part of our system, we implement a generic indexing tool for structured, semi-structured and non-structured documents. We used the vector space model (described in section II).For non-structured documents or ”Full text”, we use a test corpus CACM(Communications of the ACM) composed of 3204 documents with the title and links to bibliographic citation in the field of computer science,64 queries and a file of assessments.

We followed the steps bellow to generate the inverted documents file and the queries inverted file:

- Tokenization,
- Dropping common terms (Stop words),
- Stemming and lemmatization.

As shown in figure 4, we have an indexing process applied to corpus and a set of queries in order to have two inverted files (for the corpus and the queries), we have then a module of similarity having as input the inverted files, as output a similarity file. The result obtained with the assessment file is used by TREC-EVAL to generate a result file.

We used the TREC-EVAL result for the curve of Recall/Precision in order to evaluate our indexing approach with metrics of evaluation (Recall/Precision). Bellow some user interfaces of our indexing tool.

The first user interface (Fig. 5) allows to him to choose the language (English or French) and the type of the documents (structured, semi-structured, and non-structured).

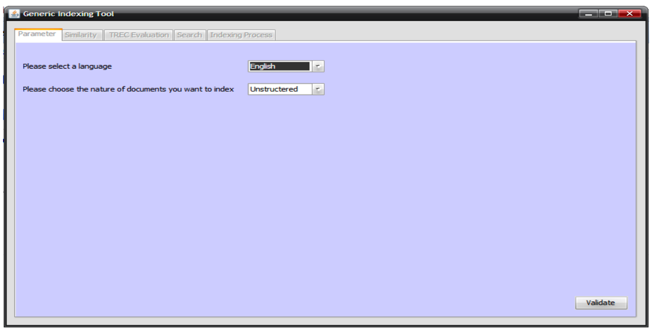

Figure 5: First User Interface

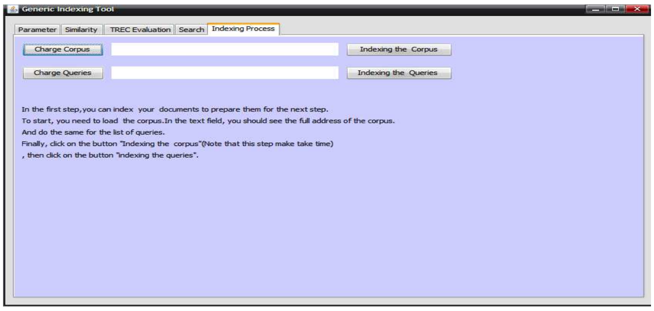

Figure 6: Second User Interface

In the second user interface (Fig. 6), the user has to load the corpus and the queries to prepare them for the indexing process.

Figure 7: Third User Interface

The third user interface (Fig. 7) is reserved for calculating the similarity for each document of the corpus with each query.

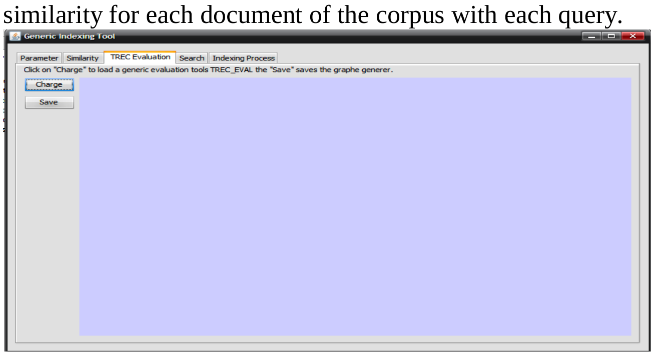

Figure 8: Fourth User Interface

The fourth user interface(Figure8) is reserved for evaluation using TREC-EVAL, the user has to load the assessment file, to save and to obtain the result of Precision/Recall(Figure 9).

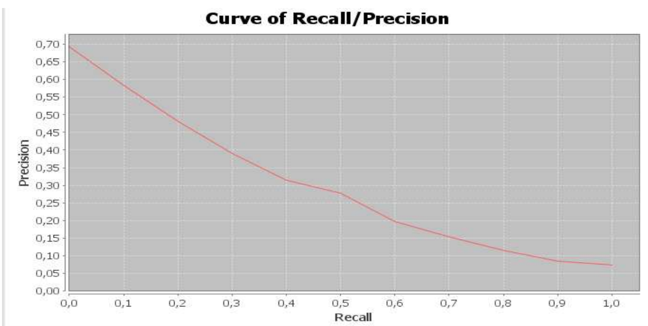

Figure 9: Curve of Recall/Precision

VII. CONCLUSION

Although a considerable number of works focused on information retrieval, some important challenges for the research community still remain. This paper provides a brief survey of the evolution of information retrieval from the classic to the contextual, the problems faced. We have proposed architecture of personalization by implying the user with representing his interests and profile into an ontology of interest, then using this ontology for re-indexing of the contents. We have exposed our indexing tool for baseline system and observed the results of Recall/Precision. As future work, we have to focus on the process of Re-indexing by implying the user (profile models, profile construction and profile evolution.).In fact, a key part of a personalization system is the user model. We have to develop an expressive model in order to encompass all aspects of the user taking into account the environment that the user interacts with and try to identify changes in user interests and needs. Finally we can say that the ultimate goal of personalization system being the satisfaction of the user. To reach this goal, the user has to be implied in the construction process in order to add the semantic value to the information retrieval.

REFERENCES

[1] C. Van Rijsbergen. A non-classical logic for information retrieval. The computer journal, 29(6), 1986,481.485.

[2] Baeza-Yates, R. and B. Ribeiro-Neto. Modern Information Retrieval. Addison-Wesley: New Jersey, 2000.

[3] S. Robertson, S. Walker, and M. Beaulieu. Okapi at trec-7: Automatic ad hoc, ltering, vlc and interactive track, 1999.

[4] Paijmans, Hans. “ SMART: Tutorial for Beginners” . September 19, 2005:

<http://pi0959.kub.nl:2080/Paai/Onderw/Smart/hands.html>.

[5] C. Bourne and B. Anderson. Dialog labworkbook. PaloAlto, Californie, USA, Second edition, Looked Information Systems, 1979.

[6] M. Lalmas, A. MacFarlane, S. M. R¨uger, A. Tombros, T. Tsikrika, and A. Yavlinsky, editors. Advances in Information Retrieval, 28th European Conference on IR Research, ECIR 2006, London, Proceedings, volume 3936 of Lecture Notes in Computer Science. Springer,UK, April 10-12, 2006.

[7] J. Wen, N. Lao, and W. Y. Ma. Probabilistic model for contextual retrieval. In Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, Shef_eld, United Kingdom,August **2004**, pages 57.63

[8] D. Grangier, A. Vinciarelli, H. Bourlard, Information retrieval on noisy text, september 2003.

[9] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. Knowledge Engeneering review, 2003, 18(2) :95.145.

[10 J. Lafferty et C. Zhai. Probabilistic relevance models based on document and query generation., volume 13. Kluwer Academic, 2003.

[11] J. Trajkova and S. Gauch. Improving ontology-based user pro_les. In Proceedings of the 8th Conference of Recherche d'Information Assistée par Ordinateur, University of Avignon (Vaucluse), France, April 26-28 **2004,** pages 380. 389.

[12] JF. Tanudjaja and L. Mui. Persona : A contextualized and personalized web search. In Proc 35th Hawaii International Conference on System Sciences, Big Island, Hawaii, January **2002,** page 53.

[13] L. Tamine, M. Boughanem, and W. N. Zemirli. Exploiting multi-evidence from multiple user's interests to personalizing information retrieval. In IEEE International Conference on Digital Information Management (ICDIM 2007), Youakim Badr, Richard Chbeir, Pit Pichappan (Eds.), IEEE Engineering Management Society., Lyon, France, octobre **2007,** pages 7.12.

[14] I. A. Research, J. Zien, J. Meyer, and J. Tomlin. Web query characteristics and their implications on search engines jason zien, j org meyer, john tomlin. In In Proceedings of the 10th International WWW Conference.Hong Kong, **2001.**

[15] A.H. Brini et M. Boughanem. Relevance feedback : introduction of partial assessments for query expansion. In Proc. of the Conference of the european Society for Fuzzy Logic and Technology, (EUSFLAT), pages, 2003, 67–72.

[16] J. Allan and al. Challenges in information retrieval and language modeling: report of a workshop held at the center for intelligent informationretrieval, SIGIR Forum ,university of massachusetts amherst, september 2002., **2003,** 37(1) :31–47.

[17] Van Dijk J.A.G.M., Ebbers W.E., Fennis B.M., van der Geest T.M., Loorbach N.R.,Pieterson W.J., Steehouder M.F., Taal E., de Vries P.W., Alter Ego: State of the art on user profiling. technical report, University of Twente, **2005.**

[18] Mobasher B., Data mining for personalization, In The Adaptive Web: Methods And Strategies Of Web Personalization. Brusilovsky P. and Kobsa A. and Nejdl W. (Ed.), **2007,** p.90-105.

[19] A. Sieg, B. Mobasher, R. Burke, G. Prabu, and S. Lytinen. Using concept hierarchies to enhance user queries in web-based information retrieval.In The IASTED International Conference on Artificial Intelligence and Applications. Innsbruck, Austria, 2004.

[20] J.-R. Wen, J.-Y. Nie, and H.-J. Zhang. Clustering user queries of a search engine. In WWW ’01 : Proceedings of the 10th international conference on World Wide Web, New York, NY, USA, 2001. ACM, pages 162–168

[21] P. Ingwersen and K. J¨arvelin. The TURN : Integration of Information Seeking and Retrieval in Context. SPRINGER, August 2005.

[22] I. Ruthven and M. Lalmas. A survey on the use of relevance feedback for information access systems. Knowledge Engeneering review, 2003, 18(2) :95–145

[23] R. W. White, I. Ruthven, and J. M. Jose. A study of factors affecting the utility of implicit relevance feedback. In Proceedings of the 28 th annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Marchionini, G. Moffat, A Tait, J Baeza-Yates, R Ziviani, N Eds, August 2003, pages 15–19.