

# Towards Contextualized Web Searches

Adriana S. Vivacqua<sup>1</sup>, Jano M. de Souza<sup>2</sup>

<sup>1</sup>RCT/PURO – Universidade Federal Fluminense (UFF)

<sup>2</sup>PESC/COPPE – Universidade Federal do Rio de Janeiro (UFRJ)

Rio de Janeiro, Brazil

avivacqua@vm.uff.br, jano@cos.ufrj.br

***Abstract.** This paper presents a collaborative project currently under discussion with a French partner, building on a prior successful experience with another French institution. We address the explosion of internet resources and ensuing information overload, which makes it hard for individuals to find relevant information. We believe in a move towards a more personalized web experience, and start by focusing on the search activity: searches are hardly ever done in a vacuum (there is an underlying activity) and should be contextualized with other actions the user is taking. In this project, we plan to develop techniques for personalized search, using models developed by French and Brazilian parties and creating new ones.*

## 1. Introduction

The large amount of information currently available online makes it hard for a user to identify the importance, quality and relevance of data retrieved. Data without provenance, web pages with conflicting or incomplete information and questionable sources are some of the problems a user must face in his or her search for information on the web. These issues make the user's task harder, as he or she faces numerous sources and is forced to identify the one that contains the information he or she needs (Gauch et al., 2007).

Most current search engines make little distinction between users and provide the same results for all of them, based only on the search terms used. However, searches are hardly ever executed “in a vacuum”: most searches serve an ulterior purpose, and are conducted as a part of a larger activity. As such, the most useful search results are those that serve that larger activity, and will not necessarily be the same for a different activity.

As a simple example, consider an arts student writing a paper on an analysis of the usage of the image of the Madonna in the Italian Renaissance period. An obvious initial search for the keyword “Madonna” in Google<sup>1</sup> yields no useful results for this student. All results relate to Madonna, the singer. The first related result appears only in the fifth page of search results. An improved search for “Madonna Italian” yields a relevant result in the first page (ranked fifth), but further searches would be needed to improve results. An even better query, “Madonna renaissance” yields several useful

---

<sup>1</sup> [www.google.com](http://www.google.com)

results in the first and subsequent pages. However, three attempts were necessary to arrive at the correct search terms for the query to generate useful results.

We contend that a contextualized approach will yield better results than the general approach currently adopted. Understanding the context in which the search is conducted and using it to improve the retrieval of web documents should result in more appropriate results. Our goal is to provide contextualized results. This also means that the definition of a good result dependent on the context in which the search was performed. Precision and recall will be used to evaluate results, as they are standard search measurements. Evaluation will also characterize the context in which the search was undertaken.

One previous approach (Lakshmanan, 2004) uses TFIDF and the vector space model, standard information retrieval techniques, to, given a set of documents that define the context of a search, expand the query to yield better results. We believe this approach can be improved on. The problem of producing personalized results has also been recognized by the web search giants Google and Yahoo!, who have recently developed ways to personalize search (in projects such as Yahoo! SearchPad<sup>2</sup>, Y!Q<sup>3</sup> and Google Web History<sup>4</sup>).

## 2. The Contextual Search Approach

Our approach to contextual searching is based on a prior successful collaboration with a French university: in collaboration with Prof. Jean Paul Barthès of Université Technologie de Compiègne, we developed an agent based collaborative work tool that used individual contexts to disseminate knowledge, under a *co-tutelle* agreement. This partnership resulted in a doctoral thesis, after a one year exchange period at UTC.

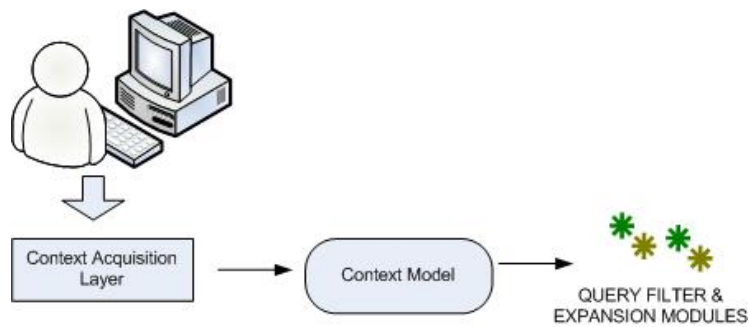
Based on this previous experienced, we propose a layered approach that captures contextual knowledge of individual work and applies it to the individual's activities. In our architecture (illustrated in Figure 1), a Context Acquisition Layer helps determine the user's activity and instantiates the contextual model for the current activity. This model is then used by query filtering and expansion modules to perform better searches. Context acquisition would ideally be performed automatically, but may involve some user input. The feasibility of automatic context acquisition will be evaluated, but is not the main thrust of this project.

---

<sup>2</sup> <http://help.yahoo.com/l/us/yahoo/search/searchpad/>

<sup>3</sup> <http://yq.search.yahoo.com/>

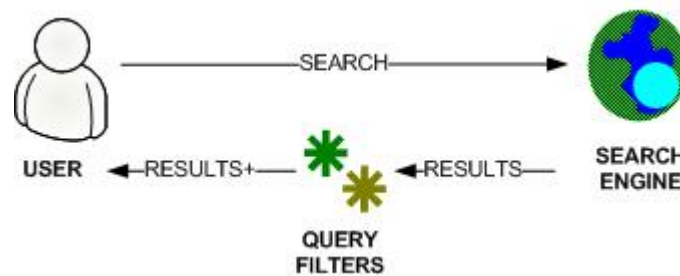
<sup>4</sup> <https://www.google.com/accounts/ServiceLogin?hl=en&continue=http://www.google.com/psearch&nui=1&service=hist>



**Figure 1. Envisioned System Architecture**

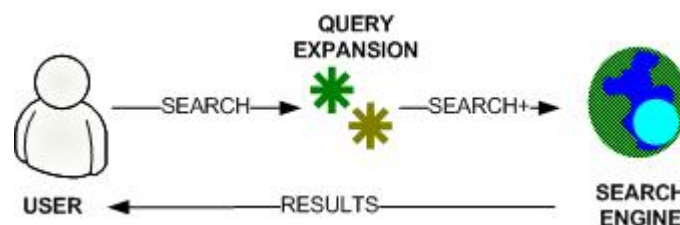
Query manipulation modules are the main focus of this project. Our goal is to improve precision and recall of web searches, by taking into account contextual information. Two lines of work will be explored (possibly in parallel):

- **Query Filtering:** this involves submitting the query, as defined by the user, to the search engine(s), and filtering the results before presentation to the user. A system would be able to analyze more results than the user normally would and rank these according to the context of the search, going beyond search engine rankings. This approach involves manipulating only query results and is illustrated in Figure 2.



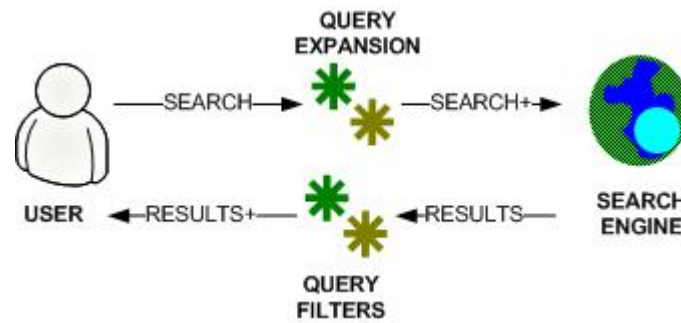
**Figure 2. Query Filtering Approach**

- **Query Expansion:** involves manipulating the query before it is submitted to the search engine, altering it through the addition of keywords or usage of advanced settings. The system would analyze the query after it is submitted by the user and alter it before submitting it to the engine. This approach is illustrated in Figure 3.



**Figure 3. Query Expansion Approach**

- **Filter and Expansion Combination:** a third approach, which combines the first two, can be tried after the first two have been developed. An expansion-filter combination might yield better results than any of the previous two alone. This is illustrated in Figure 4.



**Figure 4. Combination Approach: Query Expansion and Filtering**

These approaches will be measured individually for improvements in the results yielded. Our goal is to produce techniques to perform better web searches and later expand these to interaction with web content as a whole. The development of the Context Acquisition Layer is another topic for further work.

### 3. Discussion

We have presented an approach for the application of context in web searches. This project builds upon prior experiences with individual activity contexts developed in cooperation with a French institution. With this project, we expect not only to generate more contextually relevant search results but also to gain a better understanding of how individuals interact with information on the web and how to personalize these interactions, opening up more possibilities for research.

We are currently discussing the possibility to establish formal collaboration with a professor in France (LIP6 – Paris). This collaboration would involve using his context models as outcome for the Context Acquisition Layer and basis for reasoning on the query expansion and filtering modules. Tools already developed by his group could also be adapted as part of the Context Acquisition Layer. This would enable us to test his models and tools and explore new needs generated by this specific problem.

This project falls under the umbrella of INCT Ciência da Web (National Institute of Web Science), approved by CNPq in 2008. In addition to the professors listed above, we envision the participation of one or two graduate students in this project, which would possibly involve a post-doc or sandwich/*co-tutelle* agreement.

### References

- Gauch, S. Speretta, M., Chandramouli, A., Micarelli, A. (2007) User profiles for Personalized Information Access. In *The Adaptive Web, Lecture Notes in Computer Science 4321*. P. Brusilovsky, A. Kobsa and W. Nejdl (Eds.) pp. 54-89, 2007. Springer Verlag, Berlin Heidelberg
- Lakshmanan, H. (2004) A client side tool for contextual Web search. MSc. Thesis. Massachusetts Institute of Technology, Department of Civil and Environmental Engineering, 2004.