

Improving the Efficiency of Web Searches in Collaborative Learning Platforms

João Carlos Prates¹, Eduardo Fritzen¹, Sean W.M. Siqueira¹,
Leila C.V. de Andrade¹, and Maria Helena L.B. Braz²

¹ Department of Applied Informatics (DIA/CCET), Federal University of the State of Rio de Janeiro (UNIRIO): Av. Pasteur, 458, Urca, Rio de Janeiro, Brazil, 22290-240

² ICIST, IST, Technical University of Lisbon: Av. Rovisco Pais, Lisbon, Portugal, 1049-001
{joao.santos,eduardo.fritzen,sean,leila}@uniriotec.br,
mhb@civil.ist.utl.pt

Abstract. The web is nowadays one of the main information sources and information searching is one important issue where many advances have been reported. However, in learning environments, web searches do not consider the learning context in order to provide more adequate learning content. As collaborative learning platforms are usually web-based systems, it would be important to add a contextual web query tool to provide additional learning material from the web. This paper describes a strategy for incorporating such functionality in online social networks, as they can be used as collaborative learning platforms. In this study, the learning environment was a learning community in an online social network, in which a course was implemented. Contextual search was provided through a prototype that implements the proposed query expansion, which was evaluated using learning documents. Some results are summarized and presented.

Keywords: Social Networks, Web Search, Information retrieval, Context, Personalization.

1 Introduction

The Web is still growing on importance as people incorporate its tools and functionalities in their daily activities. The existence of so many documents available online and the need to be able to find information in such a rich repository have led to the development of search engines. A lot of research effort has been directed to obtain more efficient solutions on getting the right information from the web and search engines receive several hundred million queries each day through its various services¹. On the other hand, the disseminated use of the web has provided a whole category of online social networks.

Online social networks are another important tool that allows users to share ideas, activities, events, and interests within their individual networks. Boyd and Ellison [1]

¹ http://en.wikipedia.org/wiki/Google_Search

provided an overview of the literature related to online social networks, which were defined as: "web-based services that allow individuals to (1) build a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system."

The potential brought by these tools for human development is enormous and it is clear that the area of education can not remain indifferent to the existence of such opportunities. There are now many Web-based collaborative platforms where students and teachers interact and communicate and because many students are already using social networking sites, teachers have begun to familiarize themselves with this trend and are exploring new possibilities for increasing learning outcomes. There are now many examples of social networks and other web tools being used by teachers and students as communication tools [2], [3].

Nevertheless, when analyzing the search tools included in these environments it is possible to realize that searching for information on these platforms does not take advantage of the corresponding context in order to provide more adequate content from the web. It becomes more critical when the social networks are used as learning environment and the web searches do not consider the learning context.

This is an open issue and the work described in this paper addresses the possibility of building a contextual web query tool to be able to offer better search results. The work presented in this paper builds a strategy based on getting contextual information from shared content used in the learning platform and using a query expansion approach to get more suitable documents for the learning activities.

To be able to evaluate the value of these ideas, a prototype, based on the proposed strategy, was built and a case study was conducted in a real learning scenario at a Brazilian public university, with the aim of helping students to find educational resources in the Internet that complement their learning processes.

The remaining of this paper is organized as follows: Section 2 describes how context can be used to drive the information retrieval process. On Section 3, the architecture of the contextual query tool is presented and discussed. Section 4 presents the case study and the experiments done to evaluate the query expansion strategy and section 5 concludes presenting some final remarks.

2 Contextual Information Retrieval

In general, context is modeled through some knowledge representation. The information considered in the context can be the knowledge domain, the user profile, the business process or the user activity, the search history or even information obtained from sensors. The context model is used to make information retrieval more adequate to the user's need, thus it alters the normal process of documents retrieval. This process implies on providing new terms for expanding the original query, or removing other terms. Other possibilities for the alteration on the retrieval process are related to using the context as a basis for reclassification of the relevance order of an initial set of retrieved documents.

The context model can be created manually or automatically. In the first possibility there is direct interference of experts or users, for example through the creation of ontologies, filling out preferences forms or marking relevant documents [4], [5]. The latter provides the automatic capacity through inferences based on user behavior (analysis of clickthrough, navigation and queries history) or information from the environment [6], [7].

Although performing well in making the search tool adaptive to the context, and therefore providing more useful results, the manual context creation presents as main drawbacks the time and effort spent by users (experts or end users) in building the context model. On the other hand, the automatic creation of contexts may cause incorrect results if the inferred context does not reflect the current information needs of the user.

Another form of automatically creating the context model is through text mining, using techniques such as stemming, clustering and co-occurrence of terms. Some works that use query expansion with clustering (like the solution strategy used in this work) are presented in [5].

In advanced learning technologies, some works address the use of context in information retrieval. Zhuhadar and Nasraoui [8] capture the context with the use of domain taxonomies and user profile, and reclassify the search results according to the similarity between the terms contained in the search results and in taxonomies. Khribi et al [9] recommend educational resources based on web mining techniques. From the log files (records of visited addresses), user sessions are grouped according to similarity, and association rules are created between these sessions and visited elements. Ambrosio et al [10] use text mining techniques in selected presentations to recommend documents stored in a repository and suggest query expansion to be executed on the web.

3 The Proposed Architecture

In order to make web searching context sensitive, the tool for query expansion proposed in this work supposes that the most frequent terms in documents (links or messages) that are representative of a domain have higher probability to be in other documents that are available on the Internet and are relevant and related to this domain. Therefore, these terms are good candidates to be used to expand the queries entered by the users aiming at getting more useful results on web search engines.

As online social networks become more and more incorporated to everybody's daily tasks, they are also investigated as possible environments for other business situations. They usually provide content (media) sharing sometimes through the use of communication tools. These functionalities allow using online social networks as a collaborative learning platform. In some networks it is possible to create groups of users who share the same interest (for instance, students in a class or a course). Content sharing allows providing the necessary content material for a course, the communication tool enables exchanging ideas and therefore collaborative tasks. According to the proposal presented in this paper, through the available content it is

possible to get the domain context and, then, use this context to expand user web queries in order to obtain additional learning resources related to the subject to be learned.

For the implementation of the proposed approach, it was considered the social network facebook². The first step on creating the learning environment is to create a group, illustrated as Context Group in Fig. 1. The Learning context is acquired through published text and links.



Fig. 1. The Facebook Implementation of a Collaborative Learning Environment

For retrieving data from facebook, it is necessary to use a communication SDK/API (Open Graph API or Facebook Graph API), which is officially supported for PHP, Javascript, C#, Android and iOS. All the communication between partner app and the facebook is through the use of a JSON (JavaScript Object Notation) specification.

Some activities for preparing the text of the documents before extraction are necessary: (i) Tokenization, the process of decomposing textual content into terms; (ii) stop words removal, a list of common terms that occurs with high frequency in texts but have almost no value at all, such as prepositions, pronouns and articles; and (iii) stemming, the process of reducing a term to its root form, removing endings, affixes and thematic vowels [11].

² www.facebook.com

Considering that each document (text or link) that is representative of a context can contain different subjects, although all of them related to the context, just extracting terms based on occurrences can combine terms from different subjects in a query, thus reducing the probability of getting useful results in the web query.

Therefore it was decided to make two term extractions: one calculating the weights of the terms based only on the frequency of these terms in the whole collection of documents (in the group) in order to get the “general terms of the context”; and other in which the calculus considers the terms’ frequency in the documents of each subject that was identified in the context, the “specific terms by subject”. Each extraction of terms provides different query expansions.

In order to extract specific terms by subject it is necessary to process contextual information with information extraction activities. In order to avoid terms from different subjects being added to a query, the content of each piece of contextual information is segmented in topics, making each segment to treat a different subject and being considered an independent document in the collection.

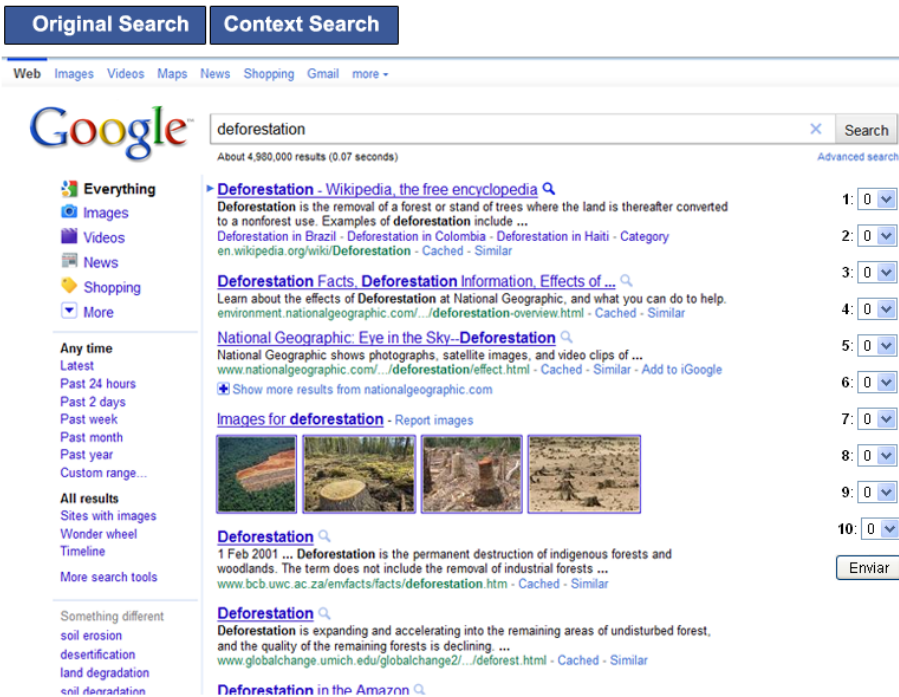


Fig. 2. The Query App’s user interface

After segmenting all the pieces of information, the segments are grouped according to their similarity, making subsets of segments. These subsets represent the different subjects that were identified in the context. For each subset, the most frequent terms are extracted.

A domain specialist must identify, for a group of selected documents, the amount of different subjects that are considered. This number defines how many groups of segments are going to be created and therefore the quantity of query expansions that will be created in the search module.

The search module gets three pieces of information from the users: the domain context, the query expression and the search engine. With these pieces of information, this module searches for the terms in the information extraction module, makes the query expansion and runs the expanded queries on the query engine.

Therefore, the students have to access the Query Application in order to make a contextualized query (Fig. 2).

4 A Case Study

To test the feasibility of these ideas a prototype was developed considering at this time only the core of the proposal and not being used in the facebook environment yet. Nevertheless it was valuable to demonstrate the applicability of the proposed approach for query expansion. The case study was conducted in a real learning scenario at a Brazilian public university, with students attending the undergraduate course "Information Systems". Context was defined as a lesson about Knowledge Management and Expert Systems. The main objective of this case study was to use the prototype to help students finding educational resources in the Internet that complement the learning process in this lesson.

The participants should express their information needs by specifying queries that represented these needs, and then the prototype with the automatic expansion facility was used to make the search. Afterwards the students evaluated the results obtained from the searches according to their relevance. In addition, after using the tool, the participants answered a questionnaire to collect data on qualitative aspects of the tool assessment.

Eighteen evaluations were collected as the class had twenty-six students, but some were allowed to work in groups. Each student or group should formulate a search expression and select the number of terms that the tool would include in the original query. They were allowed to execute the query as many times as necessary, testing a different number of terms until they felt comfortable to start the evaluation of the search results.

The results were evaluated according to three metrics proposed by different researchers and consolidated by Tang and Sun [12] for the evaluation of web search engines: first 10 full precision, search length and rank correlation.

Precision is a metric widely used in information retrieval, representing the fraction of relevant documents among all retrieved documents [11]. However, the binary judgment of relevance adopted in traditional evaluations does not take into account the different amounts of relevant information found in each document. The full precision metric tries to consider the total amount of relevant information found in the first 10 results, through the use of a five positions scale for the relevance judgments. The value "0" indicates that the result has no relevance and value "4" indicates high relevance. Therefore, the students had to judge the relevance of the first 10 results

displayed on all tabs of the search results, using a 5-point scale (as shown in the right side at Fig. 2).

The second metric was the search length, which reflects the number of not relevant documents that the user must evaluate until finding a certain number of consecutive documents considered relevant. Therefore, the lower the value, lower the effort for the user to find relevant results. As in [12], in this study the search length was defined as the number of documents evaluated until two consecutive results were found with the value of relevance greater than or equal to three. Therefore, it was possible to use data from the same evaluation tool (the 5-point scale of relevance of the first 10 results).

The last metric, called rank correlation, aims to compare the correlation between the priority that was obtained in the search with an ideal priority, where the results are sorted in descending order of relevance. The higher the correlation between the relevance of search results and the ideal prioritization, more effective is the search tool. Since the first ten results were evaluated, and the evaluation scale is five positions, the prioritization has to be considered ideal if in the first two positions it has documents with evaluation four, next two with value of evaluation equals to three, and so on until the last two results with evaluation value equals to zero.

The results are presented in percentages, representing the amount of times the result obtained in each tab was better than the others tabs.

In all metrics, the expansion with general terms of the context and the expansion with specific terms of the subjects showed better results than those obtained with the original query (Table 1).

Table 1. Results of the Metrics

Expansions / Metrics	Full Precision	Search Length	Rank Correlation
Original query	25%	16.67%	33.33%
Context expansion	66.67%	33.33%	16.67%
Subject expansion	41.67%	25%	58.33%
Any expansion	75%	41.67%	58.33%

In the full precision metric, an improvement in 75% of the cases in at least one of the expansions was observed, while the original query result only showed better results in 25% of cases. The percentage improvement was observed in the comparison of the results from the original query with the results of the expanded query with general terms of the context (66.67%), and comparing results from the original query with the results of expanded queries with specific terms of subjects (41.67%).

Improvements in results were also observed in the search length metric and in the correlation rank metric, but in smaller proportions than the full precision.

5 Final Remarks

The work presented in this paper proposes a strategy for contextual query based on shared content of a learning community implemented in an online social network. The

collaborative learning was provided through content sharing and communication tools while a web query engine was used for retrieving web documents with the learning context of the community. In order to make the query results more useful, it was considered information extraction and query expansion techniques, which were tested through a case study. The results of the experiment were presented and showed better results for the queries.

The strategy for query expansion and contextualized search presented in this paper has as main differences from related works: (i) the use of topic segmentation before clustering processes, with the objective of grouping all the different subjects found in the shared content, (ii) the use of terms of all clusters, not only the terms of the cluster most related to the query terms and (iii) quality assurance in the selection of content, consisting of educational shared resources published by the community according to its interests.

The use of activities of information extraction and query expansion can be considered to make search results more contextualized and therefore more useful to students, helping them to search for educational resources on the web, complementing the published content. The contextualization is done through the expansion of queries entered by students, adding in these queries the terms extracted from existing educational shared resources published by the community members.

A case study showed that the query expansion could be used in an educational environment to improve the educational process. Ongoing works are integrating the query module with the online social network and a case study is being conducted to evaluate the whole solution.

Some future works are: (i) analysis of the terms of the queries done by the users as a criterion for the terms expansion, for example, using as a criterion for calculating the weights of the terms the distance of each term with the terms informed in the query, in addition to frequency in the context information; (ii) improvements in the display of search results, which is tiring in cases where many expansions are showed; (iii) further analysis with bigger communities and student interactions on different communities; and (iv) providing mechanisms for reputation and rules for weighting the terms in the query expansion.

Acknowledgments. The authors would like to thank the participants of the case study. This work was partially supported by FAPERJ (through grants E-26/170028/2008 INC&T Program - Project: Brazilian Institute of Research on Web Science, and E-26/ 101.509/2010 - BBP/Bursary Representation and contextualized retrieval of learning content), CNPQ (project: 557.128/2009-9, INCT on Web Science) and by FCT Portugal, through funds of ICIST.

References

1. Boyd, D.M., Ellison, N.B.: Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication* 13, 210–230 (2008)
2. Richardson, W.: *Blogs, Wikis, Podcasts and Other Powerful Web Tools for the Classroom*, 2nd edn. Corwin Press, Thousand Oaks (2009)

3. Mora-Soto, A.: Collaborative Learning Experiences Using Social Networks. In: International Conference on Education and New Learning Technologies (EDULEARN 2009), Academia.edu (2009), http://uc3m.academia.edu/ArturoMoraSoto/Papers/114887/Collaborative_Learning_Experiences_Using_Social_Networks (last access: May 1, 2011)
4. Chanana, V., Ginige, A., Murugesan, S.: Improving information retrieval effectiveness by assigning context to documents. In: International Symposium on Information and Communication Technologies (ISICT 2004), pp. 2239–2244. IEEE Press, New York (2004)
5. Bhogal, J., Macfarlane, A., Smith, P.: A review of ontology based query expansion. *Information Processing and Management: an International Journal* 43(4), 866–886 (2007)
6. Chien, B.C., Hu, C.H., Ju, M.Y.: Intelligent Information Retrieval Applying Automatic Constructed Fuzzy Ontology. In: Proc. Conference on Machine Learning and Cybernetics, pp. 2239–2244. IEEE Press, New York (2007)
7. Wang, C., Chang, G., Wang, X., Ma, Y., Ma, H.: A User Motivation Model for Web Search Engine. In: Proc. Conf. Hybrid Intelligent Systems (HIS 2009), pp. 327–331. IEEE Press, New York (1999)
8. Zhuhadar, L., Nasraoui, O.: Semantic Information Retrieval for Personalized E-Learning. In: Proc. IEEE Conf. on Tools with Artificial Intelligence (ICTAI 2008), pp. 364–368. IEEE Press, New York (2008)
9. Khribi, M.K., Jemni, M., Nasraoui, O.: Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In: IEEE International Conference on Advanced Learning Technologies (ICALT 2008), pp. 241–245 (2008)
10. Ambrósio, A.P., Silva, L.O., Neto, V.G.: Automatic Retrieval of Complementary Learning Material for Slide Presentations. In: International Conference on Interactive Computer Aided Blended Learning, ICBL 2009 (2009), <http://www.valdemarneto.com/pdfs/artigoICBL.pdf> (last access: May 1, 2011)
11. Manning, C.D., Raghavan, P., Schütze, H.: *Introduction to Information Retrieval*. Cambridge University Press, Cambridge (2008)
12. Tang, M.C., Sun, Y.: Evaluation of Web-Based Search Engines Using User-Effort Measures. *LIBRES Research Electronic Journal* 13(2) (2003), <http://libres.curtin.edu.au/libres13n2/tang.html> (last access: May 1, 2011)