Towards a Task-based Search and Recommender Systems

Gabriele Tolomei #1*2
supervised by Salvatore Orlando #1*2 and Fabrizio Silvestri *2

#*Department of Computer Science, Ca' Foscari University of Venice*
Via Torino 155, 30172 Mestre (VE) - ITALY
[1]surname@dsi.unive.it

**ISTI-CNR*
Via Moruzzi, 1 56124 Pisa - ITALY
[2]name.surname@isti.cnr.it

***Abstract*— Nowadays, people have been increasingly interested in exploiting Web Search Engines (WSEs) not only for having access to simple Web pages, but mainly for carrying out even complex activities, namely *Web-mediated processes* (or *taskflows*). Therefore, users' information needs will become more complex, and (Web) search and recommender systems should change accordingly for dealing with this shift.**
We claim that such taskflows and their composing tasks are implicitly present in users' minds when they interact, thus, with a WSE to access the Web. Our first research challenge is thus to evaluate this belief by analyzing a very large, long-term log of queries submitted to a WSE, and associating meaningful semantic labels with the extracted tasks (i.e., clusters of task-related queries) and taskflows. This large knowledge base constitutes a good starting point for building a model of users' behaviors. The second research challenge is to devise a novel recommender system that goes beyond the simple query suggestion of modern WSEs. Our system has to exploit the knowledge base of Web-mediated processes and the learned model of users' behaviors, to generate complex insights and *task-based* suggestions to incoming users while they interact with a WSE.

I. Introduction

Recently, people have started exploiting the Web not only for having access to its huge document collection but also for accomplishing their everyday activities in a simpler way. This vision is also supported by authoritative people in the Web search community. During the DEMOFall08 Conference, there was a discussion panel on "Where the Web is Going"[1], in which Peter Norvig from *Google* and Prabhakar Raghavan from *Yahoo!* basically agreed that, rather than supporting only one search at a time, WSEs should soon focus on helping people to get a bigger task done.

So far, several works have investigated how real search intent can be devised by looking at the queries users issued to WSEs [1], [2], [3], [4]. Following this research direction, we plan to go one step forward by aiming to understand users' behaviors on a *task-based* perspective. From our preliminary study of historical data stored on a very large and long-term WSE query log, we pointed out a significative set of sample activities that many users tried to achieve by issuing some query chains. We will refer to those activities as *Web-mediated processes* or *taskflows*, which typically rely on the composition of smaller tasks. This novel way of searching the Web "*by tasks to be executed*", instead of "*by documents to be retrieved*", has to be enabled by new mechanisms that we aim to investigate. Our first research challenge is to build a knowledge base from the WSE query log, which also takes into account the semantic meaning associated with tasks (i.e., clusters of task-related queries) and taskflows. Then, this large knowledge base will help to devise a model of new users' behaviors in which users perform multiple search for accomplishing their activities. Finally, this semantic knowledge will be exploited for implementing novel and task-based recommender strategies that go beyond the simple query suggestion currently available.

The rest of this document is organized as follows: Section II outlines the proposed research steps we intend to follow. Furthermore, Section III describes a possible real life scenario where task-based search and recommendation might be useful, while Section IV shows some preliminary results obtained through the work done so far. Finally, Section V provides a summary of the whole research proposal.

II. Proposed Work

The aim of this work is to investigate and propose new Web search and recommender mechanisms "*for allowing people to get their tasks done on the Web!*"[2]. Most of the activities people traditionally performed in their everyday lives can be modeled as workflows consisting of tasks somehow composed together. In our vision, people will be increasingly interested in exploiting the Web for accomplishing those activities.

Several approaches have been proposed for analyzing how people usually search the Web by mining WSEs query logs, which record information about the *search activities* of users [5], [4], [6], [7], [8]. Thus, most works aim at understanding the real intent behind queries issued by users [1], [2], [3], [4], while others deal with the identification of users' search sessions [9], [10], [11]. Anyway, we strongly

[1]http://blogoscoped.com/archive/2008-11-06-n63.html

[2]bangalore.yahoo.com/bigthinkers/Wisdom-of-Crowds.pdf

believe that historical data stored in query logs can be helpful for proving the shift of Web usage we claimed.

We organize our work according to the following 3 orthogonal steps. First, we plan to identify small search goals from a WSE query log by using a novel clustering technique for grouping topic-related queries. Each identified cluster could represent a specific user intent aiming to achieve an *atomic* task, which potentially can be a part of a more complex taskflow. Moreover, since clusters lack of *semantic* information, we have to fill up that gap by adding meaningful labels to clusters of queries. Founded clusters of queries together with their semantic labels will constitute a valuable knowledge base of the tasks people aim to perform. Therefore, we plan to learn a model dealing with how users enact complex taskflows on the Web, i.e., how atomic tasks are composed together. Finally, as the ultimate and most important goal, we aim to investigate how taskflows or parts of them can be recommended, going one step further than current Web recommender systems of modern WSEs.

In the following, we will describe all those steps more deeply.

A. Query Clustering

The very first step of our analysis aims at finding search goals associated with users of a WSE. We start choosing the 2006 *AOL* query log as the initial data set. This is a large and long-term collection, which consists of about 20 million Web queries issued by more than 657 thousands users over 3 months. Suitable mechanisms for extracting search goals are clustering algorithms applied to the whole query session of each user. Query clustering is a quite recent research issue, and it is based on the assumption that two queries belong to the same cluster, if they are likely to be topically-related [12], [13], [14], [15], [16]. As for any other clustering problem, research has to be focused mainly on two aspects: (i) the clustering algorithm to be run on data and (ii) the *similarity metrics* used for computing the distance between queries. Many similarity metrics have been proposed for query clustering, some of them derive from traditional document clustering in information retrieval domain [17], while others exploit cross-reference between queries and user activities (i.e., relevance feedback) [18]. Recently, hybrid solutions that combine both content- and feedback-based approaches have shown to be more effective [14]. Moreover, inter-query temporal interval must be taken into account, especially when dealing with long-term historical data [19].

We started devising a *centroid-based* algorithm (i.e., *K-means*-like). However, it would be hard to estimate a good value of the input argument *K*, that is the number of clusters to be produced. So, our algorithm does not need *K* as its input argument, and its running time is $O(n^2)$, where n is the total number of the queries issued by a user within a long-term session. Currently, we are also implementing two *density-based* clustering algorithms, that is *DBSCAN* [20] and *s-NN* [21]. The advantages of using a density-based algorithm are both in terms of flexibility, and complexity: in fact, either *DBSCAN* and *s-NN* produce clusters of arbitrary shape, instead of simple balls like *K-means* does. Moreover, density-based clustering algorithms are able to deal with outliers (i.e., "noisy" queries). Finally, under particular conditions the running time of those algorithms can be $O(n \times \log n)$, which sensibly improve the overall performance and scalability when dealing with very large query logs.

All those algorithms use a novel similarity metrics, which combines query content features, relevance feedback, inter-query temporal interval, and new features based on the collaborative knowledge bases collected by *Wiktionary*[3] and *Wikipedia*[4]. Since the average number of query terms ranges between 2 and 3 words [5], we expand each query with its "*wiktionarization*" and "*wikipediation*" (whenever it is possible). Basically, we exploit the Wiktionary and Wikipedia data sources for increasing the meaningfulness of each query, trying to overcome the lack of information of the query itself. Collaborative knowledge bases like Wiktionary and Wikipedia have been recently used for increasing *semantic relatedness* between words in natural language processing domain [22], [23], [24]. However, to the best of our knowledge, these features have not yet been used for query clustering.

As a final result of query clustering we find a set of search goal clusters $\mathcal{C}_i = \{c_i^1, c_i^2, \ldots, c_i^h\}$ for each user u_i of the WSE. Moreover, each goal cluster is a sequence of not necessarily consecutive queries $c_i^j = \langle q_{i,j_1}, q_{i,j_2}, \ldots, q_{i,j_m} \rangle$ that could possibly represent a specific task-based query session. Of course, *not all* the queries issued by u_i (i.e., search goals) are thought to achieve some task. Consequently, we could expect to obtain also one or more particular "noisy" clusters for each u_i, which will represent all the query sessions that are not task-based.

B. Task Labeling

Query clustering provides an effective way of grouping together *similar* queries according to a certain metrics. Moreover, query clustering implicitly breaks the whole search session of a user in a more precise way with respect to traditional timeout-based approaches [5], [25]. Thus, while we can be quite confident that queries inside the same cluster are somehow "related", we still have to understand if a cluster really represents a task, and eventually we have to figure out its real meaning. To address that issue, a *supervised learning* algorithm (i.e., classifier), which is able to automatically assign a label to previously unseen tasks, could be used. However, it is not straightforward to assign right labels to identified groups of task-based queries. In fact, giving a too specific label to a task could make that task less "reusable". Besides, a completely general label could give too few information, making the task useless. Hence, a tradeoff is needed, and our idea is that an assigned label should help defining a sort of task *template*, which in turn could be "instantiated" and specialized from time to time. As an example, let us consider a group of

[3] `http://www.wiktionary.org/`
[4] `http://www.wikipedia.org/`

task-based queries about "renting a farm in Tuscany". If we assign the label "RENT TUSCANY FARM" to that cluster, we would probably represent a high-specific task. At the opposite side, we could add the general label "RENT", which would give not enough information about the task. So, there should be two layers of information here: (i) the task label that provides a task template, and (ii) additional data for describing the task with more details.

As for any other supervised learning approach, evaluating the outcome of a task labeling algorithm needs a corpus consisting of manually labeled tasks (i.e. *training set*). To build such a corpus, we developed a Web application that allows human evaluators to "simulate" the optimal task-based query clustering algorithm over a sample set of user search sessions. For each manually identified task, the evaluator adds a tag and optionally a longer description. Thus, we expect to collect a sample labeled data source, which will represent a semantic knowledge base of users search goals (i.e., taxonomy of tasks). Moreover, this *ground-truth* can also be useful for evaluating the results obtained with our query clustering algorithms, in terms of different measures (e.g., *precision* and *recall*).

In some way, a taxonomy of tasks would allow also to collect "groups of Web resources" related to task-based queries that together concur to achieve a specific task. Roughly, information available from this taxonomy is at a higher level of abstraction with respect to data stored on classical WSE index (i.e., group of Web resources vs. Web page). Finally, this sort of manual training phase will also be useful for the third step of our work. In fact, traditional Web recommender systems suggest items that users could potentially be interested in, by analyzing *patterns* available from stored transactional databases (i.e., *Amazon.com*, *MovieLens*, etc.). Moreover, the items users deal with are somehow "predefined" (i.e., books, music, movies, etc.). On the other hand, task-based Web transactions are definitely more free and involve many interactions with several entities, making hard to collect all data within a single transactional database of users' tasks. Thus, "items", that is tasks, which should be recommended might spread across several kind of resources (i.e., queries, Web pages, semantic tags, etc.).

C. Web-mediated Process Recommendation

Recently, popular WSEs have started to support users with query suggestions to help them in formulating better queries, hence to quickly satisfy their needs. Query recommendation techniques are based on the behaviors of users recorded in WSEs query logs [26], [10], [27], [28]. We strongly believe that the shift in Web usage should force next Web retrieval systems to make a step forward and to rely on novel recommender mechanisms [29]. We claim that next Web recommender systems should deal with two orthogonal issues: (i) *what* items should be recommended and (ii) *how* recommendation should be performed. Of course, suggestions will still take place at search-time, but they should be provided by taking into account a *window* of users' search activities, instead of recommending items on a search-by-search basis.

Regarding the first issue (i), several items dealing with smaller tasks as well as entire taskflows could be recommended. Thus, a key point here is how suggested information concerning tasks and taskflows are presented to users. As an example, items to be recommended could be the most relevant queries and/or Web resources associated with a certain task or taskflows, as well as their semantic labels.

Concerning the second issue (ii), we believe that the most suitable suggestion mechanism will be provided combining both *content-based* and *collaborative* approaches, also known as *hybrid* recommender systems, as Adomavicius et al. show in their paper [29]. Collaborative approaches make use of user profiling to provide recommendations. Typically, a *user profile* contains items together with their expressed ratings. However, in our case, items whose users are interested in are not directly available. On the other hand, items could be figured out from the semantic information associated with user's historical behaviors (i.e., groups of semantically-related queries and visited Web pages), following content-based approaches. So, content-based approaches will be useful for generating items, which in turn can be recommended in a collaborative way on the basis of users' profiles. Moreover, user profile could be either *short-* or *long-term* depending on the amount of historical data from which it was created, and it should take care of user's *contextual information* as well as her dynamic behavior.

III. Scenario

Let us suppose *Alice* wants to organize a birthday party for her baby. Typically, Alice will ask her favorite WSE issuing a query like "birthday party". Then, Alice will look at the retrieved results: some of them will be certainly "relevant" for her purpose, but mostly in a broader sense (i.e., Web pages), and they do not answer in terms of the process as a whole. Thus, Alice could rely on recommended queries that are typically more specific and partially cover the overall process. In fact, recommended queries might refer to tasks involved in organizing a birthday party (i.e., "birthday party supplies", "birthday party themes", etc.), but still without any particular correlation. So, in our opinion, WSEs together with recommender systems should organize results for *driving* Alice to the mission she wants to achieve, instead of providing "plain answers" like Web pages. In particular, recommender systems should enrich traditional search results on a task-basis perspective, namely suggesting all the small tasks composing the whole taskflow inferred by comparing between the user's query with the knowledge base of tasks (i.e., "rent a place", "order on-line a cake", "buy a gift", etc.). At the same time, WSEs should provide a new, task-oriented interface for showing retrieved results. The key point here is to "recognize" what taskflow Alice is willing to perform by simply looking at her profile. At first glance, a possible solution should take into account the short-term profile of Alice (i.e., one or a few issued queries) and compute their similarity relatedness with representative queries in the taskflows knowledge base.

IV. Preliminary Results

We believe that the outcomes of our work will affect different research topics. Regarding query clustering, we noticed encouraging results using the collaborative knowledge bases provided by Wiktionary and Wikipedia for computing the similarity between queries. Consequently, better query clustering provides a better way of splitting the entire query log into meaningful user sessions, so that each cluster will identify a search goal session more precisely.

Currently, we have almost completed the first step of our planned activities, and we are collecting the ground-truth using our task-labeling Web application, for evaluating the results of our clustering algorithms. Evaluation of identified query clusters will be expressed in terms of classical information retrieval metrics like *coverage*, *precision*, *recall*, and *F-measure*.

In the following, we will show a significative sample of clusters that have been identified with the first centroid-based algorithm we mentioned in Section II-A. In particular, Figure 1 shows three sample clusters together with a representative query for each cluster, identified for a specific AOL user.

Fig. 1. Three sample clusters for an AOL user.

Those clusters are somehow related to the taskflow for organizing a birthday party. Different colors are assigned to different atomic tasks composing the whole process (i.e., "buy a gift", "prepare a cake", "choose a theme party").

V. Conclusions and Future Work

According to many authoritative researchers in Web domain, the way in which users search the Web is rapidly changing. Thus, Web Search Engines (WSEs) may expect users to do multiple searches when trying to accomplish their everyday activities. This novel way of searching the Web "*by tasks to be executed*" instead of "*by documents to be retrieved*" has to be enabled by new mechanisms, which should be able to deal with Web-mediated processes. Hence, our first research challenge is to analyze a very large, long-term log of queries submitted to a WSE, and associating meaningful semantic labels with the extracted tasks (i.e., clusters of task-related queries) and taskflows. This large knowledge base constitutes a good starting point for building a model of users' behaviors. The second research challenge is to devise a novel recommender system that goes beyond the simple query suggestion of modern WSEs. Our system has to exploit the knowledge base of Web-mediated processes and the learned model of users' behaviors, to generate complex insights and task-based suggestions to incoming users while they interact with a WSE.

References

[1] A. Broder, "A taxonomy of web search," *SIGIR Forum*, vol. 36, no. 2, pp. 3–10, 2002.
[2] D. E. Rose and D. Levinson, "Understanding user goals in web search," in *WWW '04*. ACM, 2004, pp. 13–19.
[3] U. Lee, Z. Liu, and J. Cho, "Automatic identification of user goals in web search," in *WWW '05*. ACM, 2005, pp. 391–400.
[4] B. J. Jansen and A. Spink, "How are we searching the world wide web?: a comparison of nine search engine transaction logs," *IPM*, vol. 42, no. 1, pp. 248–263, 2006.
[5] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large web search engine query log," *SIGIR Forum*, vol. 33, no. 1, pp. 6–12, 1999.
[6] M. Richardson, "Learning about the world through long-term query logs," *ACM TWEB*, vol. 2, no. 4, pp. 1–27, 2008.
[7] F. Silvestri, R. Baraglia, C. Lucchese, S. Orlando, and R. Perego, "(query) history teaches everything, including the future," in *LA-WEB*, 2008, pp. 12–22.
[8] S. Orlando and F. Silvestri, "Mining query logs," in *ECIR*, ser. LNCS, vol. 5478. Springer, 2009, pp. 814–817.
[9] A. Spink, M. Park, B. J. Jansen, and J. Pedersen, "Multitasking during web search sessions," *IPM*, vol. 42, no. 1, pp. 264–275, 2006.
[10] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna, "The query-flow graph: model and applications," in *CIKM '08*. ACM, 2008, pp. 609–618.
[11] R. Jones and K. L. Klinkner, "Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs," in *CIKM '08*. ACM, 2008, pp. 699–708.
[12] D. Beeferman and A. Berger, "Agglomerative clustering of a search engine query log," in *KDD '00*. ACM, 2000, pp. 407–416.
[13] J. R. Wen, J. Y. Nie, and H. Zhang, "Query clustering using user logs," *ACM TOIS*, vol. 20, no. 1, pp. 59–81, 2002.
[14] L. Fu, D. H.-L. Goh, S. S.-B. Foo, and J.-C. Na, "Collaborative querying through a hybrid query clustering approach," in *ICADL*, 2003, pp. 111–122.
[15] K. W.-T. Leung, W. Ng, and D. L. Lee, "Personalized concept-based clustering of search engine queries," *IEEE TKDE*, vol. 20, no. 11, pp. 1505–1518, 2008.
[16] J. Yi and F. Maghoul, "Query clustering using click-through graph," in *WWW '09*. ACM, 2009, pp. 1055–1056.
[17] G. Salton and M. J. Mcgill, *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., 1986.
[18] G. Salton and C. Buckley, "Improving retrieval performance by relevance feedback," pp. 355–364, 1997.
[19] S. Chien and N. Immorlica, "Semantic similarity between search engine queries using temporal correlation," in *WWW '05*. ACM, 2005, pp. 2–11.
[20] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *KDD*, 1996, pp. 226–231.
[21] L. Ertöz, M. Steinbach, and V. Kumar, "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," in *SDM*, 2003.
[22] T. Zesch, C. Müller, and I. Gurevych, "Using wiktionary for computing semantic relatedness." in *AAAI*, D. Fox and C. P. Gomes, Eds. AAAI Press, 2008, pp. 861–866.
[23] T. Zesch, C. Müller, and I. Gurevych, "Extracting lexical semantic knowledge from wikipedia and wiktionary," in *LREC '08*, 2008.
[24] D. Milne and I. H. Witten, "Learning to link with wikipedia," in *CIKM '08*. ACM, 2008, pp. 509–518.
[25] D. He and D. J. Harper, "Combining evidence for automatic web session identification," in *IPM*, 2002, pp. 727–742.
[26] Z. Zhang and O. Nasraoui, "Mining search engine query logs for query recommendation," in *WWW '06*. ACM, 2006, pp. 1039–1040.
[27] Q. He, D. Jiang, Z. Liao, S. C. H. Hoi, K. Chang, E.-P. Lim, and H. Li, "Web query recommendation via sequential query prediction," in *ICDE '09*. IEEE Computer Society, 2009, pp. 1443–1454.
[28] R. Baraglia, F. Cacheda, V. Carneiro, D. Fernandez, V. Formoso, R. Perego, and F. Silvestri, "Search shortcuts: a new approach to the recommendation of queries," in *RecSys*. ACM, 2009, pp. 77–84.
[29] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions," *IEEE TKDE*, vol. 17, no. 6, pp. 734–749, 2005.