# Contextual query classification in web search

**Lynda Tamine-Lechani, Mariam Daoud, Dinh Ba Duy, Mohand Boughanem**
IRIT, University of Paul Sabatier
118 Route Narbonne, 31400 Toulouse, France

## Abstract

There has been an increasing interest in exploiting multiple sources of evidence for improving the quality of a search engine's results. User context elements like interests, preferences and intents are the main sources exploited in information retrieval approaches to better fit the user information needs. Using the user intent to improve the query specific retrieval search relies on classifying web queries into three types: informational, navigational and transactional according to the user intent. However, query type classification strategies involved are based solely on query features where the query type decision is made out of the user context represented by his search history. In this paper, we present a contextual query classification method making use of both query features and the user context defined by quality indicators of the previous query session type called the *query profile*. We define a query session as a sequence of queries of the same type. Preliminary experimental results carried out using TREC data show that our approach is promising.

## 1 Introduction

It is well known that with the increase of the information available on the Web, it is increasingly difficult for search engines to satisfy the user information need. Classical content-based systems are characterized by "one size fits all" approaches, where the IR process is based on the query-document matching pattern by considering the keyword query as the only clue that specifies the user information need. Contextual IR becomes a promising area for improving retrieval effectiveness. In [Allan and al., 2003] contextual IR is defined as follows: *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 need.* Several works in contextual IR exploit the user context that refers generally to user interests, preferences and task, in order to personalize the search [Teevan and Dumais, 2005], [Liu and Yu, 2004], [Tamine-Lechani *et al.*, 2008], [Daoud *et al.*, 2008]. Query context is usually driven by the user's intent during a search task. Specifically, web search user's intent could be driven by three main query types : (1) informational query related to the information/document finding task (what is the name of the US president) (2) navigational query related the site finding task (find the URL of the site X), (3) transactional query to the online web service finding task predicting (download a file, find music, etc.). Query type classification according to the user intent have been largely studied [Jansen *et al.*, 2007], [Rose and Levinson, 2004] in order to carry out a query specific retrieval for improving web search accuracy.

Most existing query classification strategies are based on exploiting specific query features considered as good indicators for predicting the user intent driven by the query. Several features depending on lexical query attributes, query length, usage rate of query terms in URL, etc. were distilled in [Bomhoff *et al.*, 2005] from a log file for a query classification task using machine learning algorithms. Another range of query features such as POS tags information and usage rate of query terms in anchor texts have been introduced in [Kang and Kim, 2003] to classify navigational and informational queries. An extended work was followed in [Kang, 2005] that takes into account transactional query identification by defining transactional query feature based on usage rate of query terms in service hyperlinks of the web pages. Supporting transactional queries is also treated in [Li *et al.*, 2006] where a transaction annotator allows classifying each web page as being transactional or not based on transactional terms or actions (such as download, buy, register, etc.) identified in title and anchor text of the web page. In addition to the anchor link distribution feature, user click behavior of a specific query are taken into account in [Lee *et al.*, 2005] as the main query feature used for the query classification task into navigational and informational type. User click behavior is modeled by click distributions, which captures how frequently users click on various answers and average number of clicks per query. Indeed, for a navigational query, the user is most likely to click on only one result that corresponds to the Website the user has in mind.

Once the query is classified, query specific search aims at improving the search accuracy by combining various evidences extracted from document content, links, URL and anchor information [Kang and Kim, 2003], [Westerveld *et al.*, 2001], [Li *et al.*, 2006]. Improving search accuracy for navigational and transactional query is achieved in [Jacques Savoy, 2003] by combining the document scores computed based on multiple representations. The first representation is based on the web page content. The second one is based on the anchor text and titles of all the pages pointing to the current one. The last representation is based on the texts surrounding the *TITLE* and *BIG* tags. A more specific approach for improving transactional search accuracy is proposed in [Kang, 2005] by combining the content-based document score with a transactional one computed according to the number of the available service hyperlinks

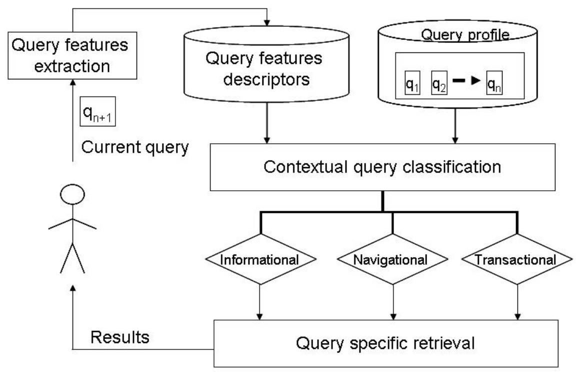


Figure 1: Contextual query classification using both query features and query profile

in the web page.
Unlike these previously cited works, we attempt to combine evidences from query features and user context defined by quality indicators of the previous query session type called the *query profile* in order to predict the user intent behind the query. Estimating the query type using the query features is based on how much the query satisfies navigational or transactional features. However, estimating the query type based on the query session type is based on the assumption that search sessions can be clustered according to the user intent [Nettleton *et al.*, 2006]. Thus the current query session type would be beneficial to better estimate the current query type. In our point of view, query features are insufficient to determine the user intent behind the query. One of the navigational query features is that query terms contain usually entity names or proper names. For instance, queries that contain verbs such "Preserve and Prevent" may not be classified as navigational. In order to deal with query type ambiguity, we assume that the query type held by the current search session is highly relevant to estimate the type of a new submitted query. While most of existing works use machine learning strategies to assign each query to one and only one type, we are aware that a query may have multiple intents so we are interested to estimate probability of classifying a query into one or more category type. We are interested in this paper to achieve contextual query classification task and we do not address the problem of query specific retrieval.
The paper is structured as follows. Section 2 presents our contextual query classification making use of evidence issued from the query features and evidence issued from the query profile. In section 3, we evaluate the effectiveness of our framework using TREC data. Conclusion is described in section 4.

## 2 Contextual query classification

As depicted in figure 1, our query classification method relies on using query descriptors extracted from query features (depending on transactional, navigational or informational query types) in addition to the query profile. Query type is used further for improving the search accuracy based on query specific retrieval approaches. Along this paper, we focus on the query classification task.
Classifying the query as estimating for each new query $q$, the probability to be classified as transactional, navigational or informational considering both query features $qf$ and query profile $qp$ defined as the current query type session. For this aim, we propose to compute, for each query $q$, the probability $p(q/qf, qp)$ to be classified as transactional, navigational or informational considering both query features $qf$ and query profile $qp$. With this in mind, we compute: $p(q/qf, qp)$ where $q$, $qf$ and $qp$ are random variables taking values in $\{I, N, T\}$, where $I$ refers to an informational query, $N$ to a navigational query and $T$ to a transactional query. Assuming that the query features are independent from the observed query session, we compute:

$$p(q/qf, qp) = p(q/qf) * p(q/qp) \quad (1)$$

### 2.1 Using evidence from query features

Here, we compute the probability $p(q/qf)$ of a query $q$ to be classified as transactional, navigational or informational using only the query features $qf$. We define the query feature vector by taking into account all query types. Feature vector $F(Q)$ of query $Q$ is defined as:

$$F(Q) = (\|Q\|, Vb, Tr, Ti, \sigma_a, \sigma_t)$$

where $\|Q\|$ is the query length (number of terms), *Vb*, *Tr* and *Ti* boolean values indicating whether or not the query contains respectively verbs, transactional terms (such as download, buy, etc.) and interrogative terms, $\sigma_a$ (resp. $\sigma_t$) is the usage rate of query terms in anchor text (resp. in page title) in the document collection, computed as follows:

$$\sigma_a(Q) = \frac{\sum_{t_i \in Q} \frac{nA_i}{n_i}}{\|Q\|}, \sigma_t(Q) = \frac{\sum_{t_i \in Q} \frac{nT_i}{n_i}}{\|Q\|}$$

$nA_i$ (resp. $nT_i$) is the number of documents containing term $t_i$ in anchor text (resp. in page title), $n_i$ is the number of documents in the collection containing term $t_i$. For each query, we build a navigational feature descriptor $QF_N(Q)$ and a transactional feature descriptor $QF_T(Q)$ as Boolean vectors defined as follows:

1. $QF_N(Q) = (CF_1^{(N)}, CF_2^{(N)}, .., CF_6^{(N)})$, where
   $CF_1^{(N)} = 1$ if $\sigma_a(Q) \geq \sigma_a$, 0 otherwise
   $CF_2^{(N)} = 1$ if $\sigma_t(Q) \geq \sigma_t$, 0 otherwise
   $CF_3^{(N)} = 1$ if $\|Q\| \leq 3$, 0 otherwise
   $CF_4^{(N)} = 1$ if Vb=F, 0 otherwise
   $CF_5^{(N)} = 1$ if Ti=F, 0 otherwise
   $CF_6^{(N)} = 1$ if Tr=F, 0 otherwise

2. $QF_T(Q) = (CF_1^{(T)}, CF_2^{(T)}, .., CF_6^{(T)})$, where
   $CF_1^{(T)} = 1$ if $\sigma_a(Q) \leq \sigma_a$, 0 otherwise
   $CF_2^{(T)} = 1$ if $\sigma_t(Q) \leq \sigma_t$, 0 otherwise
   $CF_3^{(T)} = 1$ if $\|Q\| \geq 3$, 0 otherwise
   $CF_4^{(T)} = 1$ if Vb=V, 0 otherwise
   $CF_5^{(T)} = 1$ if Ti=V, 0 otherwise
   $CF_6^{(T)} = 1$ if Tr=V, 0 otherwise

We notice that the first two features in the feature descriptors $QF_N(Q)$ and $QF_T(Q)$ reflect the fact that navigational queries have usage rates in anchor text and page titles higher than those of transactional ones. The third feature is the length property, which reflects the fact that navigational queries tend to be short compared to the transactional ones. The remaining features reflect lexical query constraints (Vb, Ti) or specific transactional property $Tr$ on the query terms that specifies the transactional query type. Let $c_i$ (resp. $c_j$) be a random variable corresponding to category $C_i$ (resp. $C_j$) taking values in $\{I, N, T\}$. We compute the probability that a query $q$ belongs to a category

$C_i$ under the condition of its classification considering the query features $qf$, as follows:
$p(q = c_i/qf = c_j) =$

$$\begin{cases} \alpha_i * p(qf = c_j) \ if \ c_i, c_j \in \{N, T\} \\ 1 - \sum_{c_j \in \{N,T\}} p(qf = c_j) \ otherwise \end{cases} \quad (2)$$

where

$$p(qf = c_j) = \frac{\sum_{i=1..6} CF_i^{(c_j)}}{\|QF_{c_j}(Q)\|}$$

$\alpha_i$ is a decay factor depending on the classification performance of category $C_i$. The higher the number of satisfied conditional features is, higher is the probability that query $q$ belongs to category $C_j$. We consider above, that informational category is the default query category.

## 2.2 Using evidence from the query profile

At this level, we estimate the probability $p(q/qp)$ of query $q$ to be classified as transactional, navigational or informational using only the query profile $qp$. Let $q$ be the current query. We define the query profile $QP^{(c_i)}$ as the set of previous queries of the same category $C_i$. $QP^{(c_i)} = \{Q_1, \ldots, Q_{m-1}, Q_m\}$. A key aspect of estimating the query type likelyhood using the query profile is to define an average session length $w_i$ for each query type $C_i$ and compute a probability of changes to a new query session type. We define the probability that a query $q$ belongs to category $C_i$ under the conditions that query profile is almost of category $C_j$ as:
$p(q = c_i/qp = c_j) =$

$$\begin{cases} 1 - \frac{\|QP^{(c_j)}\|}{w_j} \ if \ \left[(c_i = c_j) \wedge (QP^{(c_j)} < w_j)\right], \\ \frac{\|QP^{(c_j)}\|}{2 * w_j} \ if \ \left[(c_i \neq c_j) \wedge (QP^{(c_j)} < w_j)\right], \\ 1/3 \ otherwise \end{cases} \quad (3)$$

where $\left\|QP^{(c_j)}\right\|$ is the query profile length including queries of category $c_j$, $w_j$ is the average session length of type $c_j$ computed across the past user search sessions. $\left\|QP^{(c_j)}\right\| / w_j$ is the probability that the query is classified as a new query type $c_i$. This estimation relies on the assumption that when the number of queries $\left\|QP^{(c_j)}\right\|$ of the same category $c_j$ tend to be equal to the average session length $w_j$ of type $c_j$, the probability that the user changes his search intent category becomes high. Hence, since we have three query types, the probability of change from a user intent $c_j$ to a new one $c_i$ is equally distributed to each of the two possible user intents, namely $\left\|QP^{(c_j)}\right\| / 2 * w_j$. We consider furthermore, that equal probability of $1/3$ is assigned to the user query intent, as being informational, navigational or transactional, when $\left\|QP^{(c_j)}\right\|$ becomes equal to $w_j$.

# 3 Experimental evaluation

In our experimental evaluation, we study the incidence of query profile, defined as the current session type (according to the user intent) in the query classification task. We evaluated our classification methods based solely on the query features *(qf)* and then our combining method using both the query features and the whole query profile *(qf and qp)*. Our results are then compared to those obtained using TiMBL [Daelemans *et al.*, 2000], a classification based on a supervised learning software package.

## 3.1 Experimental setup and data set

We used three query sets issued from TREC data test (table 1) in our experiments. We divided each query set into two query subsets. The first one is used for training our system on the values of anchor text ($\sigma_a$) and title ($\sigma_t$) threshold values using the web collection of WT10g, containing 4500 documents indexed under *Mercure* [Boughanem *et al.*, 2003]. The second one is used for testing the performance of our methods. We compare our results with those obtained by TiMBL.
TiMBL is a learning software package used in supervised classification task. Given the training query set, TiMBL constructs a classifier for each query category by storing a set of query feature descriptors (F(Q)) defined in section 2.1. Then, given a testing query set as input, the classifier infers the query category from those of the most similar feature descriptors in memory. By training the classifier with one corpus and then demonstrating its performance on another, we establish its robustness. In order to measure

Table 1: Query data set tests description

| Query type | Collection | Training | Test |
|---|---|---|---|
| *Informational* | TREC 10 | 451-500 | 501-550 |
| *Navigational* | TREC 10 | 1-100 | 101-145 |
| *Transactional* | TREC 9 | 20001-200100 | 20101-20150 |

the query classification accuracy, precision $P(C)$ and recall $R(C)$ are computed for each query type $C$ according to the following equations:

$$P(C) = \frac{\#correct\ classification(C)}{\#total\ trial}$$

$$R(C) = \frac{\#correct\ classification(C)}{\#queries(C)}$$

where $\#total\ trial$ is the total number of test queries of different types, $\#queries(C)$ is the total number of test queries of type $C$.

## 3.2 Experimental results

In our query features-based classification, we used training query sets to fix respectively the anchor-text threshold $\sigma_a$ ($\sigma_a$= 0.39) and the page title threshold $\sigma_t$ ($\sigma_t$=0.19) used to test our classifier. Then, we classified testing queries according to $Max_{q \in \{I,N,T\}} p(q/qf)$ defined in formula (2). Moreover, we test our combining query classification by creating a query sequence from the testing set, where we alternate among the three query categories (I, N, T). We supposed that the average length of the query profile is set to 5, 3 and 5 respectively for informational, navigational and transactional category. Indeed, informational and transactional sessions tend to include high number of queries comparatively to navigational sessions to satisfy the user information need. Then, we classify a query according to $Max_{q \in \{I,N,T\}} p(q/qf, qp)$ defined in formula (1).
Table 2 shows the classification accuracy results of informational (Info), navigational (Nav) and transactional (Trans) test queries. We can see that our classification method using solely query features gives slightly better accuracy than TiMBL method, especially for informational and navigational test queries. When using our combining query classification method, we see that informational

Table 2: Test query classification results

| | Info | | Nav | | Trans | |
|---|---|---|---|---|---|---|
| *Measure* | *P* | *R* | *P* | *R* | *P* | *R* |
| Timbl (%) | 4.14 | 12 | 29.66 | 95.56 | 24.83 | 72 |
| qf (%) | 4.83 | 14 | 30.34 | 97.78 | 21.38 | 62 |
| qf+qp (%) | 15.79 | 42 | 24.06 | 96.97 | 23.31 | 62 |

query classification accuracy produces higher precision and recall compared to TiMBL query classification method and query features based classification method. This can be explained by the fact that query test classification in TiMBL depends on the selected training query set. Indeed, when the test queries have feature descriptors far from the training query ones, TiMBL shows less performance in detecting true query type. However, our query classification method relies on estimating probabilities for each possible query type based on both query specific feature descriptor constraints and the query profile.

Evaluation results prove that integrating the query profile in the query classification task allows predicting more accurately the query category, especially the informational one. We expect to define more additional specific features of all query types in order to improve their classification accuracy.

## 4 Conclusion and outlook

In this paper, we presented a contextual query classification which relies on estimating probability of classifying a query according to the user intent type defined as navigational, transactional or informational. Probabilities are computed based on using evidences issued from both the query features and the query profile defined by the current query type session. Experiments show that classification results based on combining the query features and the query profile outperforms those obtained using the evidences separately and also TimBl classification results.

In our future work, we plan to further pursue the experimental evaluation by defining additional transactional features in order to improve transactional query classification accuracy. Moreover, we attempt to achieve a query specific retrieval that returns more targeted web results so that to improve search accuracy of each query type.

## References


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

[Bomhoff *et al.*, 2005] M.J. Bomhoff, T.W.C. Huibers, and P.E. van der Vet. User intentions in information retrieval. In R. van Zwol, editor, *DIR2005*, pages 47–54, 2005.

[Boughanem *et al.*, 2003] Mohand Boughanem, Karen Sauvagnat, and Cecile Laffaire. Mercure at trec 2003 web track - topic distillation task. In *TREC 2003: The Twelfth Text Retrieval Conference*, pages 343–348, 2003.

[Daelemans *et al.*, 2000] W. Daelemans, J. Zavrel, K. Slot, and A. Bosh. Timbl: Tilburg memory based learner, version 3.0, reference guide. In *ILK Technical report*, 2000.

[Daoud *et al.*, 2008] Mariam Daoud, Lynda Tamine-Lechani, and Mohand Boughanem. Using a concept-based user context for search personalization. In *Proceedings of the International ICDMKE Conference. To appear*, London, U.K., 2008.

[Jacques Savoy, 2003] Yves Rasolovo Jacques Savoy. Hyperliens et recherche d'information sur le web 1001. In *JADT'04: proceedings of JADT 2004*, pages 325–333, New York, NY, USA, 2003. ACM.

[Jansen *et al.*, 2007] Bernard J. Jansen, Danielle L. Booth, and Amanda Spink. Determining the user intent of web search engine queries. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1149–1150, New York, NY, USA, 2007. ACM.

[Kang and Kim, 2003] In-Ho Kang and GilChang Kim. Query type classification for web document retrieval. In *SIGIR '03: Proceedings of the 26th annual international ACM SIGIR conference*, pages 64–71, New York, NY, USA, 2003. ACM.

[Kang, 2005] In-Ho Kang. Transactional query identification in web search. In *AIRS'05: Proceedings Information Retrieval Technology, Second Asia Information Retrieval Symposium, Jeju Island, Korea*, pages 221–232, 2005.

[Lee *et al.*, 2005] Uichin Lee, Zhenyu Liu, and Junghoo Cho. Automatic identification of user goals in web search. In *WWW '05: Proceedings of the 14th international conference on World Wide Web*, pages 391–400, New York, NY, USA, 2005. ACM.

[Li *et al.*, 2006] Yunyao Li, Rajasekar Krishnamurthy, Shivakumar Vaithyanathan, and H. V. Jagadish. Getting work done on the web: supporting transactional queries. pages 557–564, New York, NY, USA, 2006. ACM.

[Liu and Yu, 2004] F. Liu and C. Yu. Personalized web search for improving retrieval effectiveness. *IEEE Transactions on knowledge and data engineering*, 16(1):28–40, 2004.

[Nettleton *et al.*, 2006] David F. Nettleton, Liliana Calderón-Benavides, and Ricardo A. Baeza-Yates. Analysis of web search engine query session and clicked documents. In *WEBKDD*, pages 207–226, 2006.

[Rose and Levinson, 2004] Daniel E. Rose and Danny Levinson. Understanding user goals in web search. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 13–19. ACM Press, 2004.

[Tamine-Lechani *et al.*, 2008] Lynda Tamine-Lechani, Mohand Boughanem, and Nesrine Zemirli. Personalized document ranking: exploiting evidence from multiple user interests for profiling and retrieval. to appear. *In Journal of Digital Information Management*, 2008.

[Teevan and Dumais, 2005] J. Teevan and S.T. Dumais. Personalizing search via automated analysis of interests and activities. In *Proceedings of the* $28_{th}$ *International SIGIR conference*, pages 449–456, 2005.

[Westerveld *et al.*, 2001] T. Westerveld, W. Kraaij, and D. Hiemstra. Retrieving web pages using content, links, urls and anchors. In *Proceedings of the Tenth Text REtrieval Conference (TREC)*, pages 663–672, 2001.