

# A Web Search Analysis Considering the Intention behind Queries


Marcelo Mendoza

## Cite this paper

Downloaded from [Academia.edu](#) 

[Get the citation in MLA, APA, or Chicago styles](#)

## Related papers

[Download a PDF Pack](#) of the best related papers 



[Building decision trees to identify the intent of a user query](#)

Marcelo Mendoza, Juan F Zamora

[Query Intent Detection Based on Query Log Mining](#)

Marcelo Mendoza, Juan F Zamora

[Identifying the Intent of a User Query Using Support Vector Machines](#)

Marcelo Mendoza, Juan F Zamora

# A web search analysis considering the intention behind queries

Marcelo Mendoza  
Department of Computer Science  
Universidad de Valparaíso, Chile  
marcelo.mendoza@uv.cl

Ricardo Baeza-Yates  
Yahoo! Research  
Barcelona, Spain & Santiago, Chile  
ricardo@baeza.cl

## Abstract

*The identification of the intention of the users behind the queries can be useful to improve the precision of the list of documents recommended by the Web search engines. That is why, recent works have focused themselves in the construction of query classifiers following the categories proposed in the scientific literature. These works have based on query representations using two sources of main information: text and click-through data. Despite of the before mentioned we have little understanding about the nature and behaviour of the variables used to characterize queries. In this work we analyse the behaviour of the variables looking for a way to improve their comprehension and to identify the characteristics that exactly allow that the query classifiers improve their precision. The analysis shows that the variables based on text have a better performance in the discrimination of the categories than the ones based on click-through data. Among these variables, the query length (number of terms that compound a query), the Levenshtein distance between snippets and queries, and the PageRank metric are recommendable features to work with query type classifiers.*

## 1 Introduction

The web search engines are nowadays the most used information retrieval systems around the world. Over the big data base that is the Web, the users request to find pages that are highly relevant to their queries. Studies based on logs of commercial search engines [4, 11, 2] show that the queries have in average among two or three terms, and the collection of terms of the queries is scattered. Focused on the problem to determine the meaning of the query, the use of the terms entail several limitations, for instance given that the collection of terms is scattered, it is difficult to find relationships among them. Moreover, the terms do not always provide the meaning of the query. All of these entail that the ranking functions based on terms are limited in precision.

To tackle this problem, the attention of the web community has focused on the identification of the intention behind the query. In a first approach categories or kind of queries have proposed according to the intention of the users. Among these works Broder [3] and later Rose and Levinson [10] have consolidated a taxonomy of queries. Works based on these taxonomies have tackle the problem of building query classifiers based on different sources of information, between them the terms of documents and queries [5, 1], click-through data [8] or combinations of both sources [6]. The main goal of all these efforts have lied in effectively stating the belonging of a query to a particular category that allows the search engine to use an adequate ranking algorithm to this category improving the precision of the list of recommended documents.

### 1.1 Contributions

Despite the fact that there are an amount of important works about this problem, we understand a little about the behaviour of the variables that allow to characterize the queries according to a given categorization. This work is focused on the review of the most common variables used in the categorization of queries, based mainly on the study of two sources of information: the terms of documents and queries, and the click-through data. The purpose of this work is to illustrate the behaviour of these variables based on the query taxonomies proposed in the scientific literature, that allow to determine which of these variables are effectively more useful in a classification process.

### 1.2 Related Work

The definition of an adequate query taxonomy in the Web had its first approach in the work of Broder [3], who proposed 3 kind of queries according with the user's intentions: informational, navigational and transactional. Broder understood as informational queries the ones that the user formulates when she/he need to find accurate information related with a specific topic. The intention of the user in

this kind of queries was mainly focused on the reading of this information. The navigational queries belonged to the ones that the user formulates when her/his intention was finding a particular site. The intention behind the query was basically to navigate through the site. Finally the transactional queries belonged to the ones where the user's intentions were to make some transactions in the found site, as for example to download some resources, purchase something, among other uses. Starting from an experiment based on the opinions of experts, a set of queries were classified using the proposed categories. As a result the navigational, informational and transactional categories were distributed in a 20%, 50% and 30%, respectively.

Broder's taxonomy was later refined by Rose and Levinson [10] having in mind the interaction of the users with the selected resources. Basically the three categories proposed by Broder were maintained refining the informational and transactional types. Finally they got a total of 11 categories. Starting from experiments based on the opinions of experts, the queries were classified in the proposed categories. Just following the results given by Broder's categories, distributions that give a larger rate (60% app.) to the informational type and a minor rate (12%) to the navigational type were obtained.

Once consolidated the taxonomies proposed by Broder and refined by Rose and Levinson the construction of query classifiers had taken the effort of several works. Kang and Kim [5] proposed the construction of classifiers characterizing the queries according to the distribution of the terms that composed them. For doing that, over a set of queries classified by experts, they got 2 collections of terms frequently used in the query formulation of the type (class-kind) informational and navigational. Using measures of mutual information between both collections and characteristics as the distance among the terms of a query and the terms in the titles and snippets of the selected documents, they can determine if the term of a query is of general use or if it belongs to an informational or navigational query. The classifier based on this kind of evidence has precision of 80% app.. A disadvantage of the proposed method is that it does not consider polisemic queries. Moreover, most of the queries are constituted by few terms. Finally the method is limited in range to the manual classification that generates the initial collections that allow to build a classifier.

Lee *et al.* [6] proposed to study the click-through data to characterize queries according to the Broder taxonomy. The idea is to observe the distribution of the position in the ranking of the selected documents. Intuitively, an informational query should concentrate more clicks on the postponed parts of a ranking being different of the navigational queries, that concentrate their clicks on the first positions of the ranking. Using binary classifiers and combining these evidence with sources of information based on text, Lee *et*

*al.* built a query classifier for the informational and navigational categories. Using a set of queries manually classified by a group of experts, they evaluate the precision of the classifier getting a 90% app.. It is auspiciously comparable with the Kang and Kim method. Although the work allow to motivate the utility of the click-through data for the classification of queries, it does not allow to state conclusive asseverations due to the limitation in range of the experiments based on the opinions of the users.

A similar approach was used by Liu *et al.* [8] to build a query classifier. They proposed 2 metrics based on click-through data that allow to characterize queries: nRS (number of query sessions that registered clicks before a  $n$  position given by the ranking) and nCS (number of query sessions that registered minus of  $n$  clicks). Using a decision tree, they got a near 80% precision over a set of manually classified queries. Similar results were gotten by Baeza-Yates *et al.* [1]. The authors propose to represent queries using the terms of the documents selected in the query sessions registered in the query log files. Using the TF-IDF diagram, the problem of classification over a set of manually categorized queries by a group of experts was studied. The authors were focused in 3 categories: informational, non-informational and queries of ambiguous kind. Using a supervised approach based on SVM, accuracies near 80%, 60% and 35% were obtained for the informational, non-informational and ambiguous categories with levels of recall of 0.9, 0.4 and 0.2, respectively. The authors concluded that the use of terms of documents allows to improve the representation of the queries, being possible to identify interesting relationships in the results. The study of new representations and other classifications techniques to improve the behaviour of the classifiers was proposed by the authors.

### 1.3 Outline

The remaining of the work is organized as follows: Section 2 presents the preliminaries of the analysis, focused on the definition of some basic concepts, the presentation of the data set and the description of the query categorization process. Section 3 presents the analysis based on terms studying the relationships among the terms collection of queries and documents, the large of the queries and the relationships among the query terms and query titles and the snippets of the documents selected in the query sessions. Section 4 presents the analysis based on click-through data, checking the variables behaviour mixed up in the description of queries sessions, as well as usage or link metrics such as nRS, nCS [8] and PageRank [9]. Finally in Section 5 we show conclusions and future work.

## 2 Preliminaries

We understand by *query instance* a query formulated to a search engine in a determinate moment, following by zero or more document selections. Consequently, and following the definition introduced by Silverstein *et al.* [11], a *query session* will consist in a sequence of query instances made by the same user in a limited range of time plus the register of all the selections made by that user in this period. We excluded in this definition query sessions without selections, that we will call *empty query sessions*.

### 2.1 Query log data pre-processing

A log file register the interactions of the users with a search engine in a period of time. Among the registered interactions are query formulations, selections of documents and navigational clicks. To fulfill the goals of this work we will just recover the formulated queries and the selected documents, that means, the interaction of the users that define non-empty query sessions. From now on we will call this set of data *query log data*. We will organize the query log data in a relational data base, in order to expedite the analysis. In the keyword analysis, the vocabulary of terms of queries and documents has been processed to eliminate accents, digits and punctuation. Moreover the stop words of the collection have been eliminated. The data base was processed in a DBMS PostgreSQL 8.2 in a Athlon-XP of 2.26 GHz with 1 GB de RAM and 320 GB of hard disk, using Linux as the operative system.

### 2.2 The log data and the categorization process

We will work over data generated by a Chilean search engine called TodoCL. TodoCL mainly covers the .CL domain and some pages included in the .COM and .NET domain that are hosted by Chilean ISP providers. The data set belongs to a log file that registered the activity of the search engine for a period of six months. The file contains 127,642 queries related to 245,170 query sessions. In the file were registered 617,796 selections over a set of 238,457 different URLs. The users selected an average of 4,84 URLs per query.

A subset of the collection of queries was manually classified by a group of evaluators. This data set is composed for 6,042 queries. Due to the amount of queries, the manual classification is a slow process. To expedite the process the evaluators use a tool software that allowed them to select the corresponding categories in a more convenient way.

Following the categorization proposed in [1], the set of queries was classified into 3 categories: informational

queries, non-informational queries and queries of ambiguous intention. We understand as informational queries the queries that are formulated by the users where the intention is to find information or specific contents associated to the meaning of the query. Frequently the information will be distributed in different sites or web pages. Besides the kind of interaction with the user will be reduced to the reading of the page. It is also true that the information is in a static way in the web page, that is, it is not created as an answer for the query of the user.

As non-informational queries we understand the following: the intention of the user is to interact with the found site or web page doing some kind of transaction, that is, downloading some file, purchasing or selling some object, or finding a site, in that case the user navigate through specific contents using the found site. Frequently, the content displayed for the user will be created dynamically as a result of the interaction of the user with the site. Besides the users will have different kind of interactions: writing, downloading, navigation, among others.

Finally we classify as ambiguous queries to the ones in which the intention of the user can not be deduced directly from the query, in some cases, because the meaning of the query can be ambiguous.

After the categorization process, 3,713 queries were classified into the informational category, 1,307 in the non-informational category, and 1,022 into the ambiguous queries. The informational queries registered 5, 378 sessions in the log of non-empty queries; the non-informational queries registered 2, 586 sessions in the log of non-empty queries; the ambiguous queries registered 1,762 sessions in the log of non-empty queries, and finally the complete collection of classified queries registered 9,726 sessions in the log of non-empty queries.

Despite of the categories considered in this work are different of the ones used in the works by Rose and Levinson [10] or by Broder [3], the rate of informational queries is the same, being this kind of queries the most used.

## 3 Text analysis

One important point in the analysis is focused on determining relationships between the vocabulary that characterize the space of queries and documents. It will be logical that both terms collections have a high correlation, this is, that the terms used to formulate queries are not so different of the ones used to write documents.

Figure 1-a shows the scatter plot of the collections of terms used in queries, figure 1-b shows the scatter plot of the collections of terms used in informational queries and documents, figure 1-c shows the scatter plot of the collections of terms used in non-informational queries and terms,

and finally figure 1-d shows the scatter plot of the collections of terms used in documents.

The scatter plots show a strong correlation among the collection of terms used in documents and a weaker one for the used in queries. It is of special interest to show that the correlation between the terms of documents and queries for both categories is similar, being a little lower to the non-informational collection. The correlation factors of Pearson associated to the pairs are the following: informational /vs non-informational queries: 0.764, informational documents and queries: 0.769, non-informational documents and queries: 0.615, and finally informational /vs non-informational documents: 0.994.

Another interesting variable for this analysis is the number of terms used to formulate queries. In figure 2-a it is shown a bar plot for this variable considering informational and non-informational queries. As the graphic shows, both categories are distributed in the same queries rate with 4 or less terms. However for queries with 5 or more terms, the most used category is the informational.

Finally for each pair query-document selected we have calculated the Levenshtein distance function among the terms that compose the query and the *snippets* (the snippet is compounded by the excerpt presented with the query result, the title and the URL of the selected document). Figure 2-b shows the distribution of the Levenshtein distance function for the informational and non-informational categories. As the figure shows, the gotten distribution for the informational category has a media of 39.67, with a standard deviation of 16.47, calculated over 12,712 pairs. The distribution gotten for the non-informational category has a media of 37.6 with a standard deviation of 13.04, over 10,540 pairs.

## 4 Click data analysis

An important source for this analysis is the click-through data. To find differences and similarities among the users that formulate informational and non-informational queries, we will analyse the following variables involved in a search query: 1) Number of occurrences of the queries in a determined period, 2) Number of sessions of queries per query, 3) Number of documents selected per query, 4) Number of documents selected per session of query, 5) Number of documents selected by the position of the document in the ranking, and finally 6) Number of sessions by the position in the ranking of the last selected document (exit position). These variables will be analysed in informational and non-informational queries contrasting them in some cases, with the behaviour of queries classified as ambiguous and also with the complete collection.

First in figure 3-a we show the distribution of the number of queries according with the number of occurrences in the

log. The graphic is in a log-log format and it was generated considering informational, non-informational and ambiguous queries. As we can observe, the graphic shows that a few queries are formulated a lot of times, and most of the queries are formulated from time to time. For example nearly 20% of the queries are just formulated once and by the other hand, less than 1% are formulated more than 10 times. This behaviour is equal for all the categories considered in the experiments.

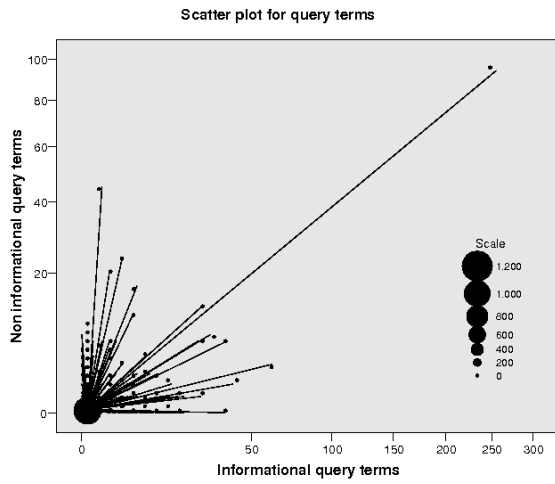
In figure 3-b we show the distribution of number of queries v/s number of sessions. The graphic shows that a great proportion of queries have a few sessions associated and a small proportion of queries is related to many sessions. This behaviour is similar for all the categories considered.

In figure 3-c we show the distribution of the number of queries by the number of clicks registered in the logs for each of them. The graphic shows that a great proportion of queries register few clicks and small proportion of queries register many clicks. For example, only the 10% of the queries register more than 10 clicks in their answers. This behaviour is similar for all the categories considered in the experiments.

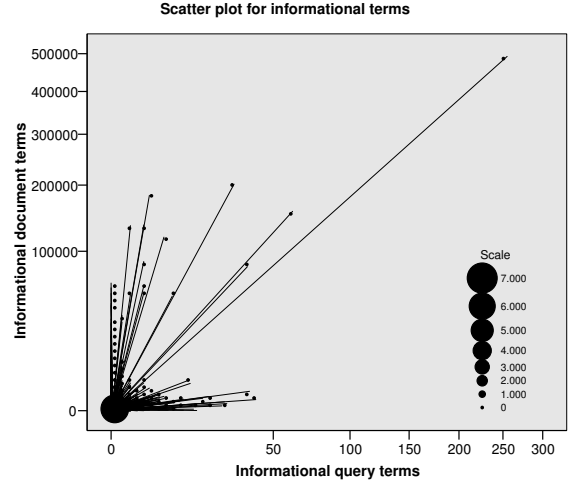
In figure 3-d we show the distribution of the number of sessions of queries v/s number of clicks by query session. The graphic is in a logarithm format in the axis of query sessions and it was generated considering informational, non-informational and ambiguous queries. The graphic shows a great proportion of query sessions registering less than 5 clicks (96% of the sessions app.). This behaviour is equal for all the categories considered. The variability observed for the number of clicks higher than 20 it is due mainly that we have less data for these stage.

In figure 3-e we show the distribution of the number of clicks v/s position in the ranking of the selected documents. For the experiment we have considered the first 30 places of the ranking. The graphic is in a logarithm format in the axis of clicks and it was generated considering queries of the informational, non-informational and ambiguous categories and the entire collection. The graphic shows that the distributions follow a power law. The distribution of the 3 considered categories do not differ in a great measure from the distribution gotten considering the entire collection and the observed variability for positions over 15 are mainly due that we have less data for these stage. Considering a power law with the form  $A \frac{1}{x^B}$ , the parameters gotten for the distribution corresponding to the entire collection are  $A = 0.198$  and  $B = 1.713$ .

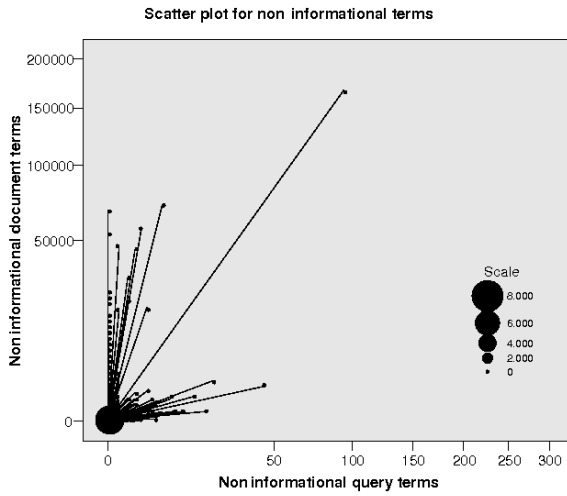
Finally in figure 3-f we show the distribution of the number of query sessions v/s the exit position in the ranking for each session of the log. The graphic is in a logarithm format in the axis of query sessions and it was generated considering queries of the informational, non-informational



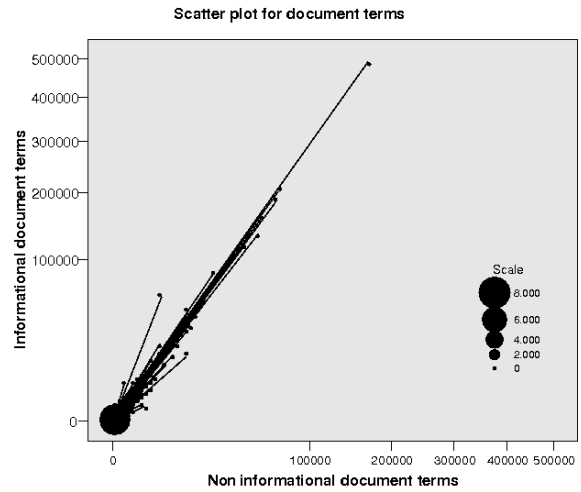
(a)



(b)



(c)



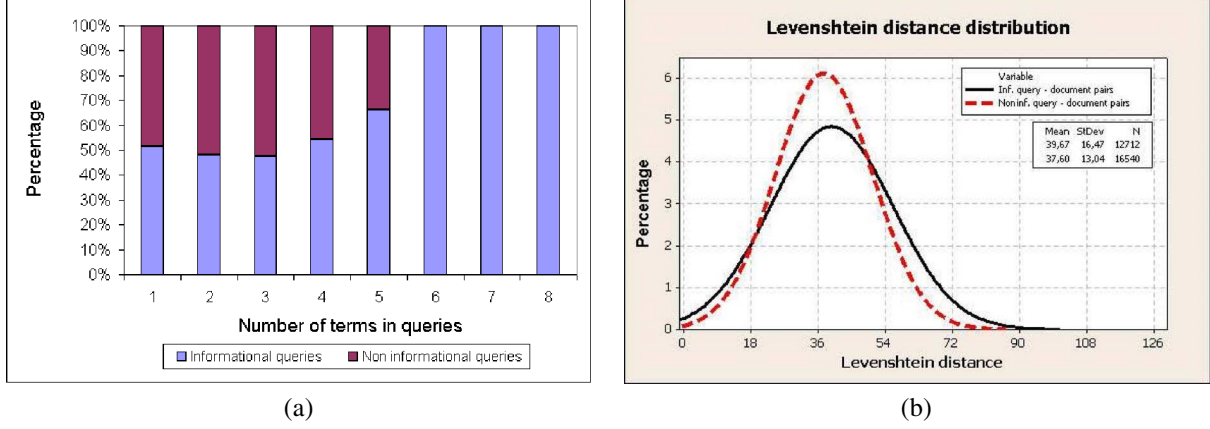
(d)

**Figure 1. Scatter plots of term collections: a) Scatter plot of query term collections, b) Scatter plot of informational term collections, c) Scatter plot of non-informational term collections, and d) Scatter plot of document term collections.**

and ambiguous categories, and the entire collection. As in the previous graphic, the distributions follow a power law and indistinctly of the considered category they do not differ of the distribution we gotten when we considered the entire collection. The peaks that can be observed in the 10, 20 and 30 positions belong to sessions where the users had checked the first, second or third page of results (the search engine shows pages with 10 results) selecting the document at the bottom of the page and leaving the session. Considering a power law in this form  $A \frac{1}{x^B}$ , the parameters gotten for the

distribution corresponding to the entire collection are  $A = 0.208$  and  $B = 1.758$ .

Other relevant variable that we should check is the duration of the query sessions. We will consider in our analysis the average of the duration time in seconds, of all the query sessions associated to queries of the same category. The sessions for informational queries have an average duration of 552.55 seconds with a standard deviation of 2181.86, over 5,378 non-empty sessions. The sessions for non-informational queries have an average duration of



**Figure 2. (a) Bar plot of the number of terms used in query formulation, for informational and non-informational queries. (b) Levenshtein distance between the query terms and the *snippets*, over the informational and non-informational query collections.**

702.28 seconds with a standard deviation of 3418.84, over 2,586 non-empty sessions. The sessions classified as ambiguous queries have an average duration of 602.26 seconds with a standard deviation of 2229.81, over 1.7628 non-empty sessions. Finally the entire collection have an average duration of 601.37 seconds with a standard deviation of 2577.48, over 9,726 non-empty sessions.

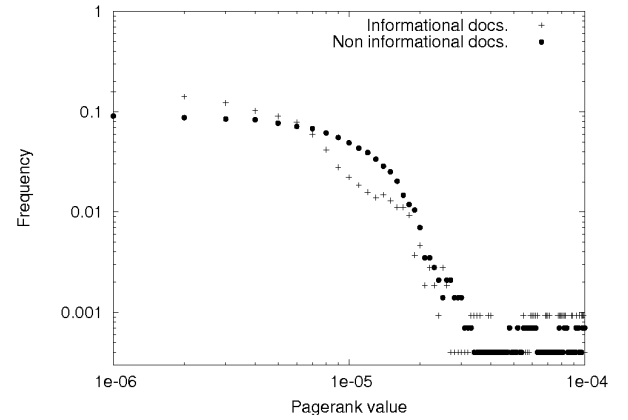
Other interesting source of information for our study is the structure of Hyperlinks of the web. We will illustrate the behaviour of the coefficient PageRank [9] for selected documents in informational and non-informational query sessions. Both distributions can be observed in figure 4. In the stage of the lower values gotten by the coefficient, this is, the interval covered among  $1e-06$  y  $5e-05$ , we can observe a higher concentration of informational documents, on the other hand, this tendency is reversed in the stage  $4e-05$  y  $9e-04$ : the Page Rank gotten by the informational category has an average of  $6.27e-05$  and in the non-informational category is  $1.52e-05$ .

Finally, we will illustrate the behaviour of the proposed metrics by Liu *et al.* [7] in our collection. The first of them belong to a number of click done (nCS) for a given query  $q$ , defined as:

$$nCS(q) = \frac{|Sess_{n,q}|}{|Sess_q|}, \quad (1)$$

where  $Sess_{n,q}$  represents the set of sessions of  $q$  that registered less than  $n$  selections and  $Sess_q$  that represents the set of sessions of  $q$ .

In figure 5 we can observe the distribution of the metric nCS over the informational and non-informational categories. Figure 5-a shows the calculate metric for  $n=2$  and figure 5-b for  $n=3$ . In the extremes, this is, for the de nCS

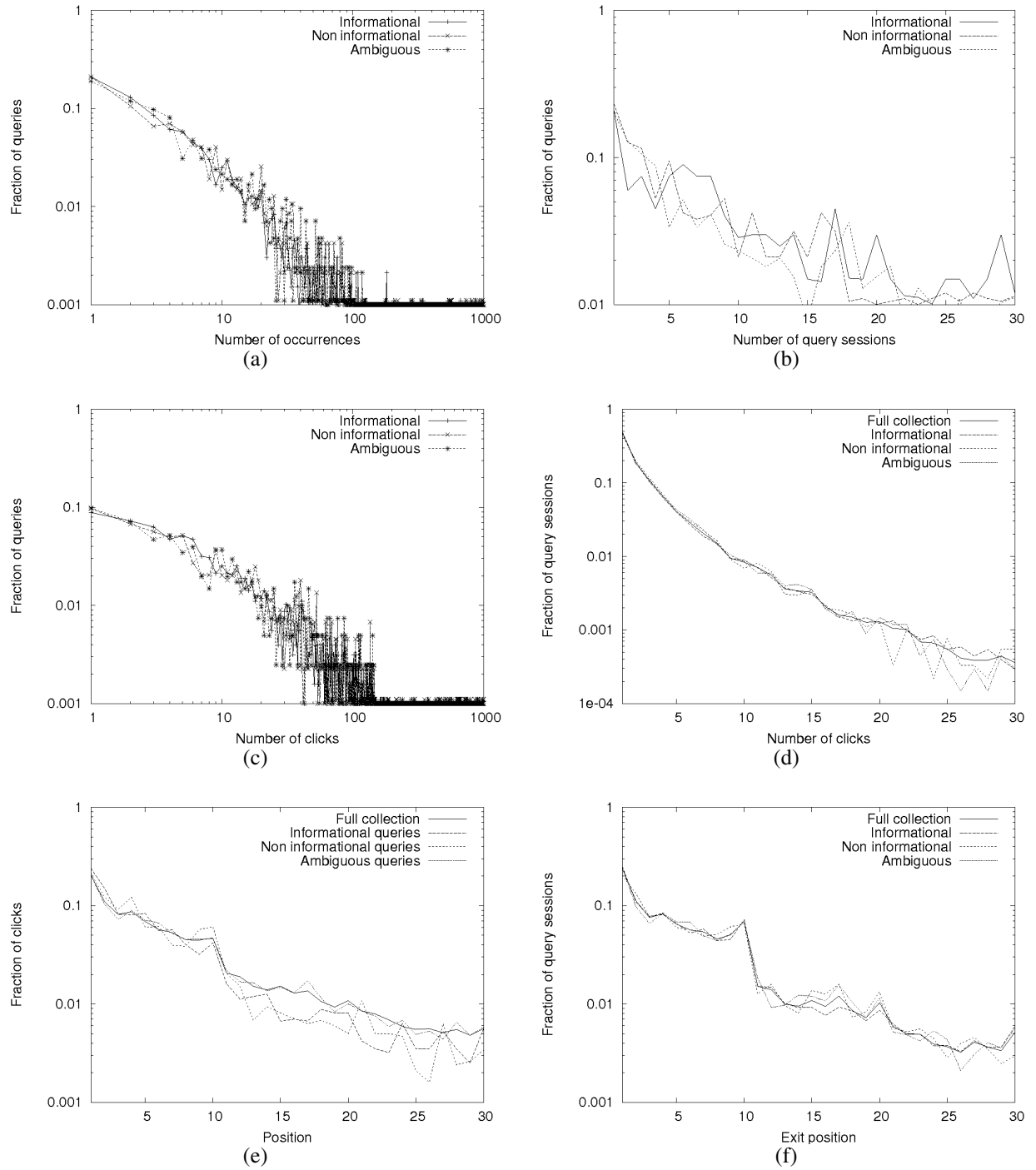


**Figure 4. PageRank distribution calculated for documents selected in informational and non-informational query sessions.**

values around the marks of 0.05 and 0.95 it is repeated for both values of  $n$  the same behaviour: the informational category exceed the non-informational category in the first interval and the tendency is reversed in the last interval. For the distribution with  $n=2$  we have an interesting point in what happen in the stage 0.05 to 0.65. In the first 3 intervals the non-informational category exceed the informational category, tendency that is reversed in the last 3 intervals. The same happens to  $n=3$  but in the stage corresponding to the category with marks 0.35 to 0.85 the tendency is reversed.

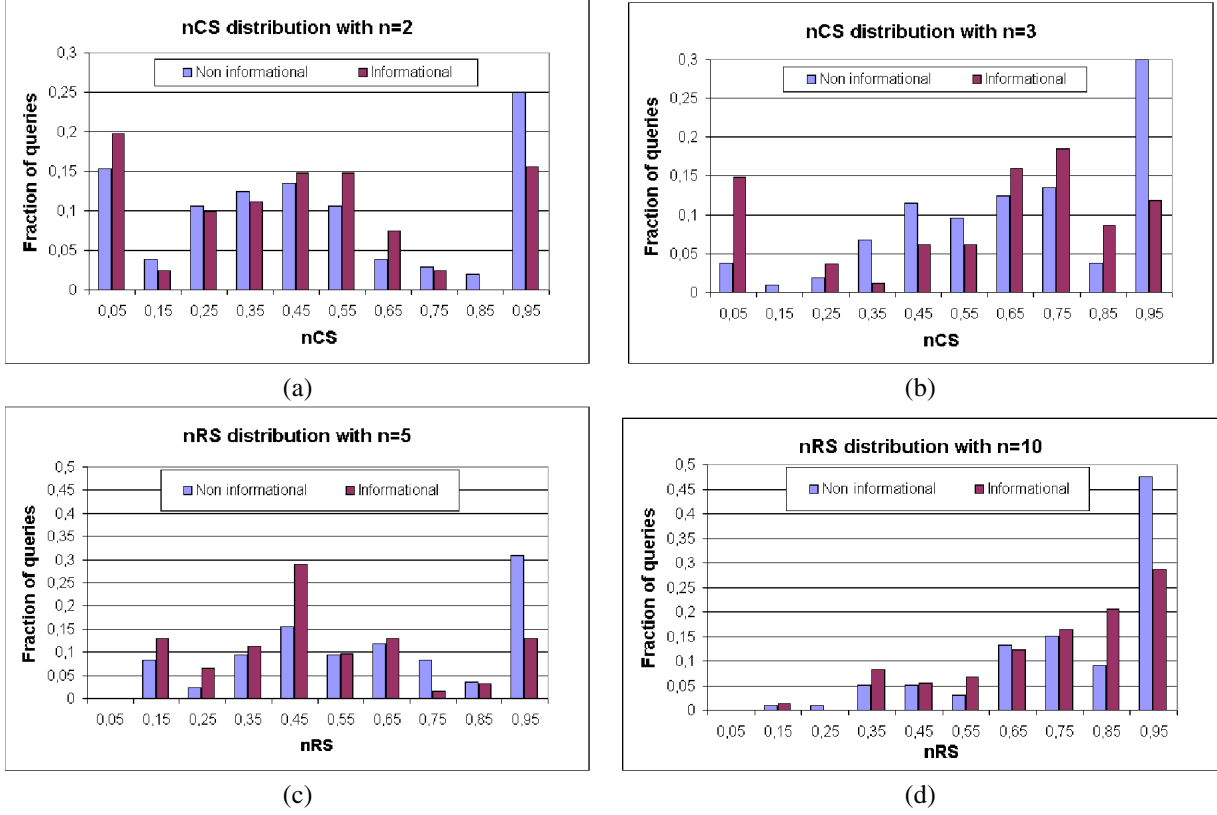
The second metric we will use is the number of revised





**Figure 3. Click data analysis by query categories:** (a) Log-log plot of fraction of queries v/s number of occurrences in the log, (b) Log plot of fraction of queries v/s number of query sessions, (c) Log-log plot of fraction of queries v/s number of clicks, (d) Log plot of fraction of query sessions v/s number of clicks, (e) Log plot of fraction of clicks v/s ranking position, and (f) Log plot of fraction of query sessions v/s exit position.





**Figure 5. Figures a) and b): NCS distribution over informational and not informational query collections: a) nCS distribution with n=2, and b) nCS distribution with n=3. Figures c) and d): NRS distribution over informational and not informational query collections: c) nRS distribution with n=5, and d) nRS distribution with n=10.**

results (nRS) for a given query defined as:

$$nRS(q) = \frac{|Res_{n,q}|}{|Sess_q|} \quad (2)$$

where  $Res_{n,q}$  represents selections only in the first n results.

In figure 5 we show the distribution of the nRS metric for n=5 (figure 5-c) and for n=10 (figure 5-d). In the class with mark 0,95, the non-informational category exceeds the informational category in both distributions. For the metric calculated with n=5, and observing the interval covered by the marks with a 0,15 to 0,45 kind we can observe that the informational category exceeds the non-informational category. The same tendency could be observed in the distribution for n=10 in the interval covered for the marks with 0,35 to 0,85 kind, the exception is the 0,65 mark where the tendency is in a contrary sense.

## 5 Conclusion

The analysis shows that the variables based on text have a better performance in the discrimination of the categories than the ones based on click-through data. Among these variables, the query length (number of terms that compound a query), the Levenshtein distance between snippets and queries, and the PageRank metric are recommendable features to work with query type classifiers. In particular, for each variable analysed in this work, we can conclude the following: 1) The collection of terms used in informational queries and documents have a higher correlation compared to the presented in non-informational terms collections. However both categories present high correlation factors, 0,769 and 0,615, respectively. Even though we could state that the text will be a more interesting source for ranking functions associated to informational queries than to non-informational queries. 2) The number of  $n$  terms of the queries is an interesting characteristic that will allow to

identify informational queries for  $n \geq 6$ . 3) The comparison among queries text and *snippets* using the Levenshtein function allows to conclude that the distance is less for the non-informational category than for the informational category (averages of 37.6 y 39.6, respectively). 4) The gotten variables using click-through data that describe the user's behavior do not allow to identify clear differences in the studied categories. 5) The duration variation variable has an interesting behaviour. We could state that as an average the non-informational queries sessions have a longer duration than the informational queries sessions (702 [s] y 552 [s], respectively) 6) The metrics *nRS* and *nCS* allow to identify differences between the informational and non-informational categories, making an analysis of the metric by intervals. However, we do not have enough evidence to state that the observed behaviours constitute a pattern, due to the observed differences are not very clear or significant. 7) The gotten PageRank for selected documents in session of non-informational queries is higher than the ones in the informational queries (6.27e-05 y 1.52e-05 average). This will allow to state that the PageRank is a more interesting metric for ranking functions associated to non-informational queries than to informational queries. We propose as a future work the design of classifiers of queries based on the variables studied in this paper. We hope to find good results with these classifiers that allow us to get low classification mistakes and finally we can improve the precision of the list of documents recommended by the search engines.

## Acknowledgements

Marcelo Mendoza was supported by DIPUV 52/2007 project from the Universidad de Valparaíso, Chile.

## References

- [1] R. Baeza-Yates, L. Calderón-Benavides, and C. González-Caro. The intention behind web queries. In F. Crestani, P. Ferragina, and M. Sanderson, editors, *Proceedings of String Processing and Information Retrieval (SPIRE)*, volume 4209 of *Lecture Notes in Computer Science*, pages 98–109. Springer, 2006.
- [2] R. Baeza-Yates, C. Hurtado, M. Mendoza, and G. Dupret. Modeling user search behavior. In *Proceedings of the 3rd Latin American Web Conference (LA-WEB)*, pages 242–251, Buenos Aires, Argentina, October 2005. IEEE Computer Society Press.
- [3] A. Z. Broder. A taxonomy of web search. *ACM SIGIR Forum*, 36(2):3–10, 2002.
- [4] B. Jansen, A. Spink, J. Bateman, and T. Saracevic. Real life information retrieval: a study of user queries on the web. *ACM SIGIR Forum*, 32(1):5–17, 1998.
- [5] I. Kang and G. Kim. Query type classification for web document retrieval. In *Proceedings of ACM SIGIR '03*, pages 64–71, 2003.
- [6] U. Lee, Z. Liu, and J. 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.
- [7] Y. Liu, Y. Fu, M. Zhang, S. Ma, and L. Ru. Automatic search engine performance evaluation with click-through data analysis. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 1133–1134, New York, NY, USA, 2007. ACM.
- [8] Y. Liu, M. Zhang, L. Ru, and S. Ma. Automatic query type identification based on click through information. In H. T. Ng, M.-K. Leong, M.-Y. Kan, and D. Ji, editors, *AIRS*, volume 4182 of *Lecture Notes in Computer Science*, pages 593–600. Springer, 2006.
- [9] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. In *Proceedings of the 7th International World Wide Web Conference*, pages 161–172, Brisbane, Australia, 1998.
- [10] D. E. Rose and D. Levinson. Understanding user goals in web search. In *WWW '04: Proceedings of the 13th international conference on World Wide Web*, pages 13–19, New York, NY, USA, 2004. ACM.
- [11] C. Silverstein, M. Henzinger, H. Marais, and M. Moricz. Analysis of a very large web search engine query log. *ACM SIGIR Forum*, 33(1):6–12, 1999.