

Characterizing Web Search Queries that Match Very Few or No Results

Ismail Sengor Altingovde
L3S Research Center
Hannover, Germany
altingovde@l3s.de

Roi Blanco
Yahoo! Research
Barcelona, Spain
roi@yahoo-inc.com

B. Barla Cambazoglu
Yahoo! Research
Barcelona, Spain
barla@yahoo-inc.com

Rifat Ozcan
Bilkent University
Ankara, Turkey
rozcan@cs.bilkent.edu.tr

Erdem Sarigil
Bilkent University
Ankara, Turkey
esarigil@cs.bilkent.edu.tr

Özgür Ulusoy
Bilkent University
Ankara, Turkey
oulusoy@cs.bilkent.edu.tr

ABSTRACT

Despite the continuous efforts to improve the web search quality, a non-negligible fraction of user queries end up with very few or even no matching results in leading web search engines. In this work, we provide a detailed characterization of such queries based on an analysis of a real-life query log. Our experimental setup allows us to characterize the queries with few/no results and compare the mechanisms employed by the major search engines in handling them.

Categories and Subject Descriptors

H.3.3 [Information Storage Systems]: Information Retrieval Systems

General Terms

Design, Experimentation, Human Factors, Performance

Keywords

Web search engines, search result quality, query difficulty

1. INTRODUCTION

A non-negligible fraction of web search queries end up with very few or even no results, especially if the query contains an infrequent term (e.g., an unusual file name produced by some malware), has one or more typos, is unusually long, or seeks an unpopular web page or content in an uncommon language. Being aware of the risk that every unsatisfied information need increases the fraction of users switching to a competitor's search service, search engines attempt to handle such hard queries by different means. In this paper, we consider the hardness of a query based on the number of

matching results and focus on queries that can match very few or no results. Queries with large result sets that do not satisfy the users' information need [3] are not in the scope of our study. Although there are some recent studies on handling long queries [2], our notion of query difficulty is based on the number of matching results.

The first contribution of this paper is to identify and analyze a large number of hard queries (using the AOL query log [5]) that originally returns very few results when submitted to one of the three major search engines, namely Bing, Google, and Yahoo! (Section 2).¹ Since such queries are very likely to include spelling errors, search engines typically accompany the original results with some alternative query suggestions (e.g., with a notification such as “do/did you mean”) or even directly blend the original query results with those of the suggested queries. We discuss several aspects of these hard queries, retrieved results, and suggestions made by the search engines via both quantitative analyses and user studies on our data (Section 3). To the best of our knowledge, no previous work in the literature discusses these issues in a real and large-scale web search setting.

Next, we focus on a very specific subset of hard queries, those that could not be handled even by the above mechanisms and remain unanswered. In this paper, we refer to such queries as “no answer” queries (NAQs) and make the first attempt to characterize NAQs submitted to a web search engine (Section 4). We believe that such a characterization is important as it may fuel the research on solving these queries, eventually leading to improvements on the search quality and user satisfaction. Solving NAQs is a vital issue in today's highly competitive search market, where users frustrated with not finding the requested information may easily switch to another search service, causing losses in revenues and brand loyalty of a search engine. Indeed, recent studies report that almost half of the users switch between search engines at least once per month [6]. According to these studies, more than half of the users state the dissatisfaction from search results as the primary reason for switching. Hence, we believe that characterizing and solving NAQs can provide significant benefits to commercial web search engines [7].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CIKM'12, October 29–November 2, 2012, Maui, HI, USA.
Copyright 2012 ACM 978-1-4503-1156-4/12/10 ...\$10.00.

¹The Yahoo! web search results are powered by Bing. We discuss this issue later in Section 2.1

2. IDENTIFYING HARD QUERIES

In our work, we use the AOL query log [5] to identify and characterize hard queries. Obviously, it is not possible to retrieve the results of all unique queries in the AOL log from web interfaces of search engines due to the query limits. Also, exhaustive sampling is unnecessary as most queries would match a large number of results and not be of interest to us. Hence, we adopt the following two-step procedure.

First, we determine a candidate set of hard queries that may return very few or no results when submitted to a web search engine. Since earlier work suggests that search engine APIs process queries over an index that seems to be smaller than the full web index [4], we believe that identifying queries that return no answers from a search engine API is a good starting point. To this end, we use a dataset from a previous study [1], where 660K unique AOL queries were issued to the Yahoo! web search API in December 2010. We select queries that return no answers (around 16K queries) and resubmit them to the same API (in the summer of 2011). We observe that the number of queries without any results drops to 11K queries. Next, we issue these 11K candidate hard queries to three major search engines (Bing, Google, and Yahoo!) and retrieve the first result pages (similar to [4]). The queries are issued to the U.S. frontends, which are supposed to have the largest index.

In Table 1, for the three search engines, we report the number of queries that return k or fewer results. For some queries, search engines provide query suggestions together with the corresponding results. Here, we consider only the number of results retrieved for the original query, not for the suggestion. Since providing a comparison of the search engines is not our goal, we arbitrarily name them as A, B, and C. The results validate the two-step procedure we used to identify hard queries: a large fraction of the 11K queries submitted to the web search interfaces originally return very few or even no answers. For instance, search engines A, B, and C return less than 10 results for 33%, 66%, and 63% of the queries. Moreover, 2% to 17% of these hard queries turn out to be actual NAQs. In the light of these findings, we can safely refer to the queries in our 11K set as hard queries.

2.1 Caveats

As our hard queries are seeded with those that do not retrieve any results from the Yahoo! web search API, we might have a slight bias towards those queries that cannot be solved by Yahoo!. To investigate this issue, from the AOL query log, we randomly sampled 6K singleton queries that are not in our initial 16K queries (non-singleton queries are likely to be solved by all three search engines). We issued them to the web frontends of all three search engines and retrieved results, which shows that the ranking of search engines with respect to the percentage of NAQs is the same as in Table 1. However, as expected, the absolute numbers are much smaller. Hence, we believe that the way we construct our query set does not introduce a significant bias against any search engine. Nevertheless, verifying our results using seed sets based on other search engine’s APIs would also be a good future research direction. We are also aware that, while we conducted our experiments, the Yahoo! web search results are supposed to be provided by Bing. Yet, we prefer to have both search engines in the discussions as i) even though the overlap observed between the results is not very low, they are not completely identical (also demonstrated

Table 1: Number of queries that return k or fewer results (only the original query results are used)

k	A	B	C
0	244 (2%)	1997 (17%)	1791 (15%)
≤ 2	1129 (10%)	6377 (55%)	6368 (55%)
≤ 10	3829 (33%)	7721 (66%)	7366 (63%)
≤ 100	7394 (63%)	9089 (78%)	8960 (77%)

Table 2: Message patterns in search result pages

P	SE	Message displayed in the search engine result page
0	All	–
1	Bing	Do you mean <suggested query>.
	Google	Did you mean: <suggested query>.
	Yahoo!	Did you mean: <suggested query>.
2	Bing	No results found for <original query>.
	Google	Showing results for <suggested query>.
		Showing results for <suggested query>.
		Search instead for <original query>.
	Yahoo!	We have included <suggested query> results. Show only <original query>.
3	Bing	No results found for <original query>.
	Google	Your search – <original query> – did not match ...
	Yahoo!	We did not find results for: <original query>.

by some of the results reported in this paper), and ii) both search engines are most likely to employ different query correction mechanisms, leading to differences in certain cases.

3. HARD QUERIES WITH FEW RESULTS

While most hard queries match very few results when processed in their original form, search engines are indeed well armed to resolve some of these queries. In particular, we found that typical query correction mechanisms also improve the performance (in terms of the number of retrieved results) for a large fraction of our hard queries, as well. Having said that, we emphasize that such correction mechanisms and associated query suggestion patterns that appear in search result pages (e.g., the well-known “do/did you mean” message) are general techniques that a search engine would fire to handle any type of typos in the queries, with or without taking into account the number of results. Our comparative analysis in this section reveals how differently search engines attempt to correct this particular set of hard queries, how they overlap or differ in the approaches employed, and how successful these corrections are in terms of user satisfaction.

Query correction for hard queries. During our analysis of the retrieved result pages for the hard query set, we observed four types of result patterns adopted by all three search engines (Table 2). In case of the first pattern, simply the answer of the submitted query is shown in the result page. As the second pattern, all three search engines return some form of query correction/suggestion together with the results of the original query. The third pattern is to immediately provide the results of another, potentially related query instead of the original query. Finally, we observe a fourth pattern when no results match the query.

In Table 3, for the three search engines, we report the number of query results falling under each pattern. Remarkably, all three search engines attempt to correct the query terms for most of the hard queries, by either providing a query suggestion (i.e., pattern 1) or directly providing the suggested query’s results (pattern 2). The search engine A provides immediate answers to the majority of the hard queries (around 62%), whereas B and C essentially handle the majority of these queries via pattern 1.

The fraction of queries that result in pattern 2 is similar for all three engines. Moreover, by using pattern 2,

Table 3: Number of queries with a certain pattern, observed at each search engine (SE)

SE	Pattern 0	Pattern 1	Pattern 2	Pattern 3
A	7,267 (62%)	1,277 (11%)	2,896 (25%)	233 (2%)
B	3,519 (30%)	4,584 (39%)	3,068 (26%)	502 (4%)
C	2,771 (24%)	5,340 (46%)	3,101 (27%)	461 (4%)

Table 4: Number of queries that return k or fewer results for each pattern and search engine (the percentages are computed with respect to all queries with a given pattern and search engine)

SE	k	Pattern 0	Pattern 1	Pattern 2
A	≤ 2	544 (8%)	126 (10%)	4 (0%)
	≤ 10	1602 (22%)	435 (34%)	17 (1%)
	≤ 100	4530 (62%)	751 (59%)	196 (7%)
	≤ 1000	4530 (62%)	751 (59%)	196 (7%)
B	≤ 2	2040 (58%)	1295 (28%)	96 (3%)
	≤ 10	2785 (79%)	2094 (46%)	189 (6%)
	≤ 100	3059 (87%)	3099 (68%)	621 (20%)
	≤ 1000	3059 (87%)	3099 (68%)	621 (20%)
C	≤ 2	1496 (53%)	1605 (30%)	69 (2%)
	≤ 10	1965 (71%)	2608 (49%)	157 (5%)
	≤ 100	2233 (81%)	3747 (70%)	557 (18%)
	≤ 1000	2233 (81%)	3747 (70%)	557 (18%)

search engines B and C can substantially reduce their NAQ ratios (the pattern 3 column). A quick comparison between Tables 1 and 3 shows that using pattern 2 helped several queries in B and C that originally return no answers (further details are discussed next). Even in this case, for 2% to 4% of the hard queries, the result pages contain no results.

Number of results. While the discussion above shows that query corrections and subsequently suggested queries can reduce the percentage of NAQs, the success of the results returned after these corrections is not clear. As evaluating around 11K results for all three search engines requires significant human effort, we limit our analysis to a comparison of the number of matching results for queries with patterns 0, 1, and 2. Table 4 reports the number of queries that return k or fewer results, for every pattern and search engine pair. We observe that queries with pattern 2 match the largest number of results, whereas directly answered queries (i.e., with pattern 0) match the smallest number of results (especially, for B and C). For instance, C returns less than 10 results for 71% of queries with pattern 0, but only for 5% of queries with pattern 2. The result counts for queries with pattern 1 is usually between those of patterns 0 and 2.

We interpret the observation above as follows. When pattern 2 is shown, the search engine is rather confident in that the user intention well matches another query, which can retrieve potentially more results than the original query. In case of pattern 1, possibly there is a good query to suggest, but the original query can also match some results. Hence, the search engine includes the suggestion in the result page but prefers to present the results of the original query. Finally, for direct results, the search engine is either very confident about its results (e.g., A finds more than 1,000 results for 38% of queries with pattern 0) or it cannot find an alternative query to recommend and hence presents whatever results the original query matches (e.g., for 58% and 53% of queries with pattern 0, B and C retrieve at most two results, respectively, but cannot suggest an alternative query).

For the queries answered with pattern 2, we further obtained the results of the original query (Table 5). Our findings show that, for search engines B and C, 49% and 43% of queries answered with pattern 2 are indeed NAQs when the original query is followed (i.e., contributing to NAQ percentages in Table 1), respectively, and presenting the results of the suggested query is somewhat mandatory for these

Table 5: Number of queries with pattern 2 for which the corresponding original query return k results (the percentages are computed with respect to the corresponding values in Table 3)

k	A	B	C
0	11 (0%)	1,495 (49%)	1,330 (43%)
≤ 2	226 (8%)	2,540 (83%)	2,333 (75%)
≤ 10	1,792 (62%)	2,842 (93%)	2,793 (91%)
≤ 100	2,113 (73%)	2,931 (96%)	2,980 (96%)

Table 6: Number of results retrieved from fake result sites for each pattern and search engine

SE	Pattern 0	Pattern 1	Pattern 2
A	38%	19%	1%
B	18%	5%	1%
C	13%	5%	1%

queries. For A, the situation is different since a significant portion of these queries can retrieve some answers. In this case, we anticipate that A uses some other clues in addition to the number of results to apply pattern 2. Even for A, which usually retrieves more results than B and C, the majority of these queries match relatively fewer results than those with pattern 0 and 1 (please compare with Table 4).

Domain of the results. While Table 5 shows that the queries with pattern 2 originally return very few results, an astute reader may object that, at least for a certain percentage of queries with patterns 0 or 1, the number of results is not so few, as we claim it to be true for hard queries (see Table 4, especially for the queries answered by A). However, we observed that, for several queries, the web pages returned in the top 10 results simply include a list of the queries in the AOL log. Obviously, such a page cannot be considered as a real answer for the query. While it is impossible to determine all such domains manually, we basically inspected the domains that appear most frequently in the results of queries with patterns 0, 1 and 2. We identified the most frequent five domains that seem to be a plain compilation of the AOL queries or URIs in these queries (aolscandal.com, aolstalker.com, robtex.com, t35.com, and iwant*.info). In Table 6, we present the percentage of answers from these domains in the first result page (i.e., up to the top 10 results) of queries with patterns 0, 1 and 2 for each search engine. Apparently, a considerable fraction of results for patterns 0 and 1 come from these few domains, implying that the number of real results for these queries, if there are any, is even smaller than what seems to be retrieved by the search engines. This observation further confirms that our process for identification of hard queries successfully yields those queries that return few results.

Overlap between search engines. In Table 7, we show the number of overlapping queries among all three search engines and their pairwise combinations. For completeness, we also report the number of queries that return a pattern by only a single search engine. We observe that the highest agreement among the three search engines is for the queries having pattern 2, i.e., among 4,946 queries associated with this pattern, 1,186 are common in all three search engines (about 24%). The agreement is also reasonable for patterns 0 and 3, i.e., there is a common set of queries that are either directly answered or could not be answered by any search engine. Interestingly, the overlap is low for queries with pattern 1. In terms of pairwise relations, B and C have a higher level of consensus while A usually differs from them.

We also investigate the overlap among different search en-

Table 7: Number of overlapping queries with each pattern for each search engine (the percentages are computed with respect to the value of the corresponding columns’ union)

Search engine	Pattern 0	Pattern 1	Pattern 2	Pattern 3
$A \cap B \cap C$	1,637 (18.86%)	273 (4.11%)	1,186 (23.98%)	114 (17.14%)
$(B \cap C) \setminus A$	754 (8.68%)	3,840 (57.87%)	1,425 (28.81%)	287 (43.16%)
$(A \cap C) \setminus B$	241 (2.78%)	152 (2.29%)	158 (3.19%)	0 (0.00%)
$(A \cap B) \setminus C$	606 (6.98%)	27 (0.41%)	164 (3.32%)	8 (1.20%)
$C \setminus (A \cup B)$	139 (1.60%)	1,075 (16.20%)	332 (6.71%)	60 (9.02%)
$B \setminus (A \cup C)$	522 (6.01%)	444 (6.69%)	293 (5.92%)	83 (12.48%)
$A \setminus (B \cup C)$	4,783 (55.09%)	825 (12.43%)	1,388 (28.06%)	113 (16.99%)
$A \cup B \cup C$	8,682	6,636	4,946	665

Table 8: Number of overlapping query suggestions for queries with patterns 1 or 2 for each search engine (the percentages are computed with respect to the corresponding values in Table 7)

Search engine	Pattern 1	Pattern 2
$A \cap B \cap C$	29 (10.62%)	245 (20.66%)
$(B \cap C) \setminus A$	2834 (73.80%)	1235 (86.67%)
$(A \cap C) \setminus B$	46 (30.26%)	40 (25.32%)
$(A \cap B) \setminus C$	2 (7.41%)	16 (9.76%)

gines in terms of suggested queries (for patterns 1 and 2). As seen in Table 8, the overlap for the three engines is twice larger for queries with pattern 2 than those with pattern 1. This may imply that all engines can detect the user intent better for these queries. As before, the overlap between B and C is very high for queries with patterns 1 or 2.

Methods for generating suggestions. In patterns 1 and 2, search engines suggest an alternative to the original query. We manually inspected all queries that are answered by either one of these patterns by all three search engines (as seen in Table 7, summing up to 273+1,186 queries for patterns 1 and 2, respectively) to identify the types of modifications made on the original query to create a suggestion. We observed that a large number of queries entirely or partially include a URI. Indeed, among the 273 queries that are answered by applying pattern 1 in all three search engines, the amount of queries with a URI adds up to 71%. For pattern 2, the percentage is smaller yet significant: 52% of 1,186 queries contain a URI. Due to the frequent presence of URIs in queries, in Table 9, we present the modifications for these two types of queries (with and without URIs), separately.

According to Table 9, there are some fundamental differences between the modifications applied to queries with or without URIs. In particular, for queries without URIs, the most common modifications are adding space between the words and then correcting typos within the terms. On the other hand, only 30% of URI queries involve an obvious typo while the rest do not necessarily contain a spelling mistake in a strict sense, but they are possibly due to the poor memory of the user, who confused “com” with “biz”, or forgot the hyphen between the terms (see the examples in Table 9). For this latter class of suggestions, search engines probably use the existence of other closely similar URIs as a clue.

Suggestion quality. To investigate the accuracy of the suggestions made by the methods discussed above, we randomly created two subsets, each with 100 queries, from the queries that yielded results with pattern 1 or pattern 2 in all three search engines. We then conducted a user study with six judges. Each judge is shown the original query and the suggestions from each search engine, and is asked to decide if the suggestion makes sense or not. In Fig. 1, we show the percentage of suggestions labeled as relevant, irrelevant, and undecided by the judges (for patterns 1 and 2). The figure shows that A yields the lowest number of irrelevant results,

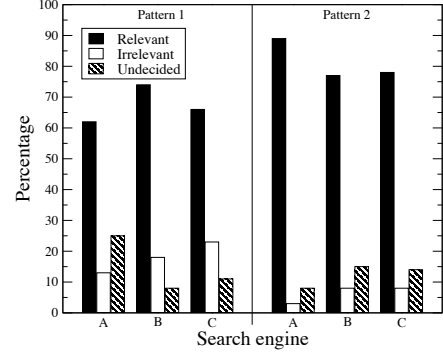


Figure 1: Suggestion quality.

but it is not the best performing search engine due to the large fraction of suggestions that are left undecided (around 25%). A closer inspection reveals that A consistently prefers to provide alternative URI suggestions, whereas the other two search engines simply split the URI into terms as a suggestion in most of the cases. This choice of A yields lots of undecided suggestions as the judges could not decide on how good the suggested URI captures the intent of the user in a number of cases. Nevertheless, for queries with pattern 1, we see that the fraction of irrelevant suggestions vary between 13% and 23%. For queries with pattern 2, all search engines provide a larger fraction of relevant suggestions compared to those for pattern 1. This is a result that further confirms the intuition that the search engines return results with pattern 2 only when they are more confident with their suggestion. In this case, the fraction of suggestions labeled as irrelevant is less than 10% for all three search engines.

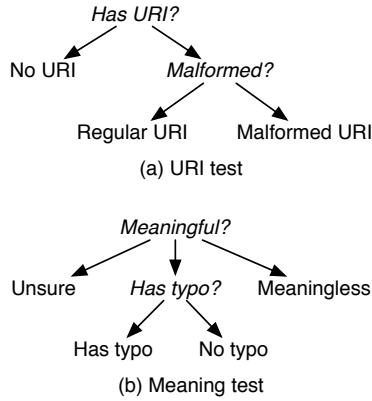
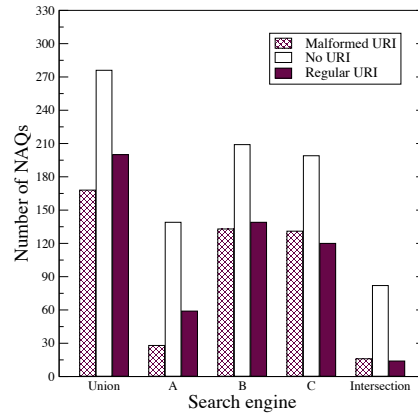
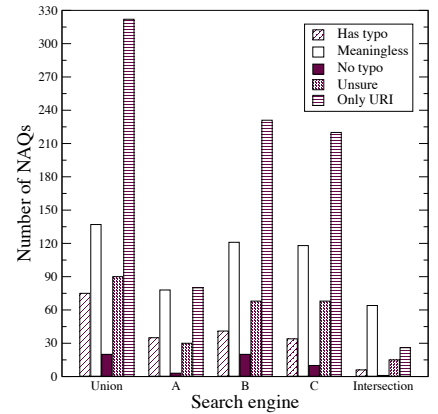
4. NO-ANSWER QUERIES (NAQS)

We now focus on queries that retrieve no answers (NAQs). In particular, we use queries that match no results in at least one of the search engines (according to Table 7, there are 665 such queries). To investigate the characteristics of these queries, we conduct a user study, where each NAQ is labeled by four human judges based on two types of tests: URI presence and meaningfulness (see Fig. 2 for details).

Our first test evaluates the presence of URIs in NAQs. Although it could be possible to automate this test via pattern matching techniques, we prefer to do it manually as it is difficult to automatically catch URIs that contain typos. The results in Fig. 3 indicate that about 57% of the NAQs contain at least one URI. About 45% of these contain at least one malformed URI while the remaining 55% are proper URIs. This shows that about one-third of NAQs aim to retrieve resources that are unknown to or not discoverable by the search engine. Hence, it may not be possible to solve these NAQs by any technique. When we compare the results across the three search engines, we observe that A is

Table 9: Most frequent modifications to queries without (M1–M5) and with a URI (M6–M11)

Modification	Original Query	Suggested query
M1 Split query string to terms	3rdgenerationgospelsingers	3rd generation gospel singers
M2 Correct typo in a term (insert/delete/replace character)	tadeair compter show	trade air computer show
M3 Combine terms in query string	cup cakes	cupcakes
M4 Add/delete punctuation	childrens hospital of birmaham	children’s hospital of birmingham
M5 Add/delete/replace term	woodiestationwagons	woody station wagons
M6 Split URI to terms	www.eldercare-today.com	elder care today
M7 Correct typo in a term in URI	www.orlandocollages.com	www.orlandocolleges.com
M8 Add/delete/replace term in URI	www.online-houses-for-sale.com	www.online-homes-for-sale.com
M9 Re-order terms in URI	ri-rvs.com	rvs-ri.com
M10 Add/delete punctuation	street-racingvideos.com	street-racing-videos.com
M11 Add/delete/replace domain extension	learn-direct-advice.co	www.learn-direct-advice.co.uk

**Figure 2: The procedure followed by the judges in the user study.****Figure 3: Distribution of NAQs based on the presence of a URI.****Figure 4: Distribution of NAQs based on the meaningfulness.**

significantly better in solving NAQs with malformed URIs. The number of such NAQs in **A** is only slightly higher than those present in the intersection set of the three search engines. Overall, the size of the intersection is much smaller than the size of the union, which implies that most NAQs with a URI are solved by at least one search engine.

Our second test is about the meaningfulness of the NAQs. If a query contains a URI, we only consider the remaining query terms. If the entire query is a URI, it is labeled as “only URI” and excluded from the test. If the meaning of a NAQ is not clear to the judge, but the NAQ has a potential to have a meaning for the user who issued it, then the judge labels the NAQ as “unsure”. NAQs that are clearly meaningless to the judge (e.g., queries that are only formed of repetitive key strokes) are labeled as “meaningless”. The remaining NAQs are considered “meaningful” and labeled as “has typo” or “no typo”, depending on the presence of a typo.

The results of this test are shown in Fig. 4. Since a considerable portion of the NAQs are labeled as “unsure”, the numbers reported for the remaining labels can act only as lower bounds. According to the results, only 3% of NAQs are meaningful and do not contain any typos. It is interesting to note that, in our study, we encountered only one such NAQ that is not solved by either search engine. At least, four out of every five NAQ that is meaningful contains a typo, which is not fixed by the spell checker. At least 21% of NAQs do not have any meaning. This final result sets an upper bound of 79% on the fraction of NAQs that a search engine can fix by employing more sophisticated techniques.

5. CONCLUSION

We provided a characterization of hard queries that match very few or no results in web search engines, which cope

with such queries by suggesting alternative queries. We conducted a user study which reveals that the majority of the suggested queries are relevant to the original query. However, there is still some room for improvement since some queries are left with no answers (NAQ). We found that half of such queries contain a URI and it is possible to develop techniques to solve 79% of NAQs, specially focusing on URIs. As a future work, we plan to further study NAQs.

6. ACKNOWLEDGMENTS

This work is partially supported by EU FP7 Project CUBRIK (contract no. 287704).

7. REFERENCES

- [1] I. Altingovde, R. Ozcan, and O. Ulusoy. Evolution of web search results within years. In *Proc. 34th Int’l ACM SIGIR Conf.*, pages 1237–1238, 2011.
- [2] N. Balasubramanian, G. Kumaran, and V. R. Carvalho. Exploring reductions for long web queries. In *Proc. 33rd Int’l ACM SIGIR Conf.*, pages 571–578, 2010.
- [3] D. Carmel, E. Yom-Tov, A. Darlow, and D. Pelleg. What makes a query difficult? In *Proc. 29th Int’l ACM SIGIR Conf.*, pages 390–397, 2006.
- [4] F. McCown and M. L. Nelson. Search engines and their public interfaces: Which apis are the most synchronized? In *Proc. 16th Int’l WWW Conf.*, pages 1197–1198, 2007.
- [5] G. Pass, A. Chowdhury, and C. Torgeson. A picture of search. In *Proc. 1st Int’l Conf. Scalable Information Systems*, page 1, 2006.
- [6] R. W. White and S. T. Dumais. Characterizing and predicting search engine switching behavior. In *Proc. 18th ACM CIKM*, pages 87–96, 2009.
- [7] H. Zaragoza, B. B. Cambazoglu, and R. Baeza-Yates. Web search solved? All result rankings the same? In *Proc. 19th ACM CIKM*, pages 529–538, 2010.