

An analysis of queries intended to search information for children

Sergio Duarte Torres
University of Twente
The Netherlands
duartes@cs.utwente.nl

Djoerd Hiemstra
University of Twente
The Netherlands
hiemstra@cs.utwente.nl

Pavel Serdyukov
Delft University
The Netherlands
p.serdyukov@tudelft.nl

ABSTRACT

Query logs contain valuable information about the behavior, interests, and preferences of the users. The analysis of this information can give insight in their interaction and search behavior. In this paper, we analyze queries and groups of queries intended to find information that is suitable for children by using a large-scale query log. The aim of the analysis is twofold: (i) To identify differences in the query space, content space, user sessions, and user click behavior. (ii) To enhance the query log by including annotations of queries, sessions and actions. The paper presents plans to use this resource for further research on information retrieval for children. We found statistically significant differences between the set of general purpose queries, and the set of children queries. We show that many of these differences are consistent with small-scale research studies in which children were observed while using web search engines.

Categories and Subject Descriptors

H.3.3 [Information Storage and Retrieval]: Query formulation, Search process

General Terms

Experimentation, Measurement

Keywords

query log analysis, children

1. INTRODUCTION

The Internet today is widely used by children for information, communication and entertainment purposes. From 2000 to 2002 the Internet access of children aged 2-17 increased from 46% to 78% in the United States [15]. Similarly, the London School of Economics reported that 75% of the 9-19 years old people in the UK have access to the Internet at home and 98% somewhere by 2004 [25]. Undoubtedly, the access and use of the Internet by children will keep increasing in these and other regions of the world in the coming years. Unfortunately, most of the current Information

Retrieval (IR) systems are designed for adults and previous studies have shown that children's information needs [32], search approaches [11] and cognitive skills [29] differ from those of adults. Thus, there is an increasing need for research aimed at understanding children's information needs and to provide IR systems that suit the characteristics of content for children.

Query logs represent valuable sources of information to understand the search process and to improve search engine systems. For instance, query logs have been widely exploited in the literature to study user's behavior/interaction with IR systems, to classify queries [12], to infer search intent [2] [10], to generate user profiles [2], to produce query suggestions [8], among others.

In this paper we explore the AOL query log [28] to compare the queries and sessions use to retrieve information for children, and queries and sessions use to retrieve general purpose information. The aim of this analysis is twofold: (i) To identify differences in the query space, user sessions and user behavior of these two types of queries. (ii) To enhance this query log by identifying children queries, sessions and actions. This resource will be used to study other important problems in IR for children as query assistance and query classification.

For the analysis we rely on the Kids and Teens section of the DMOZ directory¹ to identify the queries employed to retrieve content for children. The aim of this DMOZ section is to provide child friendly and safe content to cover the specific needs of people under the age of 18. We consider that using this directory to identify children-content queries is reasonable and realistic enough given that the content of the directory is frequently regulated and maintained by senior editorial staff, which guarantees that websites with harmful or unsuitable content for children are excluded.

Note that although it is not possible to establish if these queries were performed by children, we are still able to study the characteristics of the queries and sessions when the underlying information need is related to content for children. Moreover, it has recently been reported (from a survey of 2131 families in UK) [27] that children aged 5-15 are frequently trusted to search the Web on their own (68% and 84% for children aged 5-7 and 8-15, respectively), which suggests that there is a high chance that queries retrieving information for children are actually submitted by children.

Another important characteristic of DMOZ is the use of age tags which can be used to distinguish content suitable for kids up to 12 (kids), 15 (teens) and 18 years old (mature

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.

IIIX 2010, August 18–21, 2010, New Brunswick, New Jersey, USA.
Copyright 2010 ACM 978-1-4503-0247-0/10/08 ...\$10.00.

¹http://www.dmoz.org/Kids_and_Teens/

teens)². This feature allow us to compare the characteristics of the queries and sessions for these three target groups. In Section 2 we introduce the most relevant related work on query log analysis and children search behavior. In Section 3 we describe the data employed in this study and the methodology followed for its analysis. Section 4 present the results obtained for the children queries and sessions. In Section 5 we discuss some directions and ideas of how to improve information retrieval for children based on our findings and children search behavior. Finally in Section 6 conclusions are drawn and directions for future work are stated.

2. RELATED WORK

Although we are unaware of any previous query log analysis in IR for children, several studies have been carried out to analysis large-scale query logs of commercial search engines. Silverstein et al. [30] analyzed a query log of the Altavista search engine that contains approximately 1 billion entries. They presented an analysis of individual queries, query duplication, query sessions and correlations between query terms based on a set of descriptive measures such as query length, query frequency, session length and term frequency. According to this study users tend to utilize short queries (mean of 2.3 words per query) and user sessions are short on average (2 queries per session). They also reported that queries are not changed often by users and that 77.5% of the queries are unique, which suggests a wide variety of information needs and several ways to express them. Similar results regarding query length and query characteristics are reported in the analysis done by Spink et al. [31] on a smaller query log of the Excite search engine.

Pass et al. [28] analyze various aspects of an AOL query log such as query formulation patterns, search engine efficiency, user demographics and user’s interactions. They described the query space as vast, topically diverse and in constant change. Interestingly, they also found that 20% of the users perform approximately 70% of the queries and that less than 1% of the web domains present in the log account more than 50% of the clicks of the users. Further analysis on this query log is carried out by Brenes et al. [9] by grouping the queries and sessions based on the query popularity. They found different behaviors throughout the groups (i.e. navigational coefficient, query length, temporal length) which suggests that more fine-grained analysis are required to study query logs.

A crucial aspect to study query logs is the definition of the user session. A session is a sequence of queries issued by the user to satisfy an information need. In [8] [22] [18] the sessions are defined using a timeout cutoff between queries, which establishes that two queries are in the same session if the time difference in which they were issued is smaller than a given threshold value. Jones et al [22] showed the limitations of this method to capture individual search tasks. They proposed an automatic procedure to segment further user’s sessions in goals and missions. Goals are defined as an atomic information needs and missions as a set of related information needs. In our analysis we use a similar approach to detect goals in the query log.

In this paper we will refer to the case studies on children search behavior carried out by Bilal [5, 6, 7] and Druin [13] to compare and analyze the findings on user’s behavior drawn

²<http://www.dmoz.org/guidelines/kguidelines/>

from the query log. We will also refer to Ofcom’s study which presents an overview of media literacy among UK children from 5 to 15 years old based on a survey performed from 2007 to 2009 on 2131 families.

3. CHILDREN QUERIES AND SESSIONS

The AOL query log employed contains approximately 36 million entries and it was collected during two months in 2006. Currently, this is the largest and most up-to-date freely accessible query log in the Web Each entry contains an anonymous user ID, time of submission, rank position and domain of the URL clicked. It is important to mention that there is controversy about the usage of this query log in the research community given the privacy issues that arose by the identification of actual users in the press [24]. Nonetheless, all the results presented in this paper are obtained from accumulated counts and no actual user identification is performed. The identification of the queries employed to retrieve content for children was performed by matching the entries listed in the DMOZ kids & teens directory, which contained 45.635 entries, with the domains clicked in the entries of the query log. Given that the query log does not include the entire URL visited, matches are restricted to the cases in which only the domain is listed as DMOZ entry. Three data sets of query log entries were constructed by employing the matching procedure described above on the DMOZ entries tagged for kids, teens and mature teens.

Sessions are constructed by grouping contiguous queries submitted with a time difference smaller than t_θ and that are from the same user. A formal definition of session is shown in Equation 1.

$$S = \langle \langle q_{i_1}, u_{i_1}, t_{i_1} \rangle, \dots, \langle q_{i_k}, u_{i_k}, t_{i_k} \rangle \rangle \quad (1)$$

where $u_{i_1} = \dots = u_{i_k}$, $t_{i_1} \leq \dots \leq t_{i_k}$ and $t_{i_{j+1}} - t_{i_j} \leq t_\theta$ for all $j = 1, 2, \dots, k - 1$. The parameter t_θ was set to 30 minutes because it is the most common value employed in the literature [22] [8] [18]. We consider that this time window is also suitable for sessions expressing children information needs because it has been shown that the average time children spend to fulfill an information need varies between 10 and 16 minutes [7].

The identification of atomic information needs or *goals* is performed by grouping any pair of queries from the same user according the conditions expressed in Equation 2.

$$\left\{ q_i, q_j : \left[\begin{array}{l} t(q_i) < t(q_j) \wedge \\ ((Levenshtein_{distance}(q_i, q_j) < \alpha) \vee \\ (Word_{distance}(q_i, q_j) < \beta) \vee \\ Domain(q_i) = Domain(q_j)) \end{array} \right] \right\} \rightarrow \{0, 1\} \quad (2)$$

where α and β are set to 1. This equation groups a pair of queries if the *edit distance* is below a threshold α , the queries contain only β different words (*word distance*), or the queries are used to click in the same domain. The time restriction specifies that the queries grouped keep the order of submission. The goals are constructed by merging all the pairs of related queries in a transitive fashion. We employed the word and edit distance to find related queries because these features have been shown to be highly effective in the identification of goal boundaries [22]. The sessions and goals used to satisfy children’s information needs are those that visit at least one DMOZ domain. Three sets of sessions and

Table 1: Size of the query sets

	Queries	Uniq. Q.	Sessions	Goals
Kids	485,861	10,252	21,009	32,292
Teens	411,474	4,169	7,930	14,503
M.Teens	516,570	10,057	15,519	26,600
All log	36,389,577	10,154,747	10,769,830	8,005,597

Table 2: Frequency of query types

	All	Kids	Teens	M_Teens
By q. structure				
Question queries	2.7%	3.9%	3.7%	3.9%
Phrasal queries	56%	65.9%	64.7%	55.7%
By q. intent				
Informational	51.7%	33.5%	51.05%	49%
Navigational	21.5%	13%	32.7%	33.2%
Transactional	26.7%	53.3%	16.2%	17.5%

goals were built using these criteria. Table 1 summarizes the characteristics of the set of sessions and goals collected.

4. EXPERIMENTATION AND RESULTS

4.1 Query level results

In this section we explore the characteristics of the query entries. We considered the query length (which is measured by the number of words per query), the type of queries (question queries, phrasal queries, query intent), cue words, the rank position of the domains clicked, the query frequency distribution and the distribution of users across the datasets.

4.1.1 Query length Analysis

Query length is an indicator of the complexity of the query and the difficulty of the user to express information needs using keywords. The average query length found for the kids, teens and mature teens datasets were 3.8, 3.4 and 3.2, respectively. These values differ from the mean of the entire query log, which is 2.5 words per query, with statistical significance using the Wilcoxon signed-rank test at the 95% confidence level. The average query length of the entire data is also in line with the average length reported for other large scale query logs [30, 4].

Interestingly, these results confirmed previous studies in the field of Human-Computer Interaction (HCI). Druin et al. [13] stated that kids of age 8 to 12 tend to formulate longer queries as a consequence of their preference to write queries using natural language constructions instead of keywords, specially when the information need requires multiple phases to be solved. It has also been found that children tend to express complex information needs by directly typing the question they have [7].

4.1.2 Query type Analysis

Given these observations the queries were classified into question and phrasal queries. Question queries (e.g. *what is the leprechaun*) are defined as the queries that contain at least one of the following tokens: *what, where, why, who, when, whose, will, am, are, is, have, has, whether, which and whom* [4]. None of the queries in the query log contained the question mark character (possible due to query

normalization before the release of this query log). Phrasal queries (e.g. *words that start with the N letter*) refers to the presence of noun and/or verb phrases in the queries. The frequency of these type of queries are summarized in Table 2. In this table is shown that question queries are more frequent in the children queries (queries use to retrieve information for children) than in the general purpose query set, which is in line with Druin’s observations [13]. These findings suggest that query reduction and query segmentation techniques can be particularly beneficial for children content queries since it has been shown that longer queries are less efficient [23]. Similarly query reformulation techniques based on morphological and syntactical features of the queries can improve the search process by mapping phrases to concepts and keywords [14].

4.1.3 Query intent Analysis

Queries were also analyzed using Broder [10] classification which captures three types of user intent: informational, navigational and transactional queries. Informational queries are used to address an information need by locating content relevant to the topic of interest (e.g. areas in africa giraffes live in). Navigational queries are used to locate a specific Website, which can be the main Website of an organization or a hub site (i.e. bobthebuilder.com). Transactional queries are used to locate a Website with the aim of obtaining a product. The product may refer to an item to be purchase, an application to be executed on-line (i.e. alphabet coloring pages) or a multimedia resource to be downloaded.

The results summarized in Table 2 were obtained by manually classifying the queries using the guidelines given by Jansen et al. [20] to classify query intent (Broder categories are also used in this study). The queries were classified by sampling randomly at 15% the unique queries of the kids, teens and mature teens data set. For the whole data set a random sample of 1400 queries was obtained. This size is comparable to the size of the sample employed in a previous study on a large query log [10].

We found that informational queries are preferred in the whole, teens and mature teens data set over transactional and navigational queries. Previous studies have also found this behavior on large query logs. For instance Broder [10] reported on a random sample of 1000 queries from the Altavista log that 48% of the queries were informational, 30% transactional and 20% navigational, which are comparable to the percentages obtained in our sample.

Interestingly, this trend was not observed for the kids queries in which transactional queries are preferred (increase of 20% in respect to the average user of the query log). We found that transactional queries are mainly used in the kids and teens queries to interact with web applications (e.g. flash/java games, academic quizzes) or to obtain free on-line resources (e.g. poems, songs lyrics, coloring pages).

Conclusions in the same line have recently been drawn in Ofcom’s studies [27]. They reported that gaming is the preferred Internet activity for children aged 5-7 and second preferred for children aged 8-11. In particular 37% of the children aged 5-7 (52% for children aged 8-11) were found to use the Internet at least once per week for gaming while 19% use it for information purposes (46% for children aged 8-11). On the other hand for users aged 12 -15 informational and social activities are more popular than gaming.

Table 3: Cue clusters

	Cue word	Content words
kids	coloring	pages,free, page, easter, book, day, bible, disney, pring, preschool, butterfly, mother's, animal...
	craft	ideas, tissue, projects, kite, luau, invitation, frame moms, folded, gallon,plastic, shells, canoe...
	poems	funny, teachers, short, haiku, silverstein, concrete, shel, seasons, appreciation, mother, acrostic...
	help	thank, neopets, abc's, fluoride, teeth, insects, text, myspace, ways, magnesium, hw, guild...
	game	mystery, files, dressup, match, pool, board, sites, play, create, yahtzee, maze, math...
teens	ps2	cheats, games, game, godfather, codes, 2006, andreas, baseball, mlb, 2k6, naruto...
	videos	drift, jupiter, randy, snakes, orton, stunt, hero, racer, angel, d1...
	facts	fun, pearl, harbor,abstinence, planets, aids, jazz, yeast, thailand...
	school	statesmen, greenbriar, rule, freshman, houses, centerpieces, jones...
	life	wonderful, important, called, teenagers, george, factor, survival, character, roaring...
m. teens	body	muscles, odor, stay, infections, flexible, water, piercings, development, bruises, collagen...
	news	fox, science, channel, nasa, bahrain,latest,soft, knoxville, superstring, singel, dinosaur, iraq...
	download	orion, beholder, tv, shows, boggle, dominoes, fmv, craps, snood, knight, scrabble, collapse...
	map	middle, east, israel, egypt, okinawa, galilee, sea, jerusalem, palestine, detailed, surrounding...
	college	grants, scholarships, sinclair, community, search, darton, kissing, binge, drinking, financial...

In this case 66% use the Internet at least once in the week for informational purposes and 48% for gaming.

The lower use of informational search in the kids queries compared to the other query sets can be caused by the current lack of specialized IR applications to satisfy children's information needs or to the unsuitability of most of the content in the Web for these users. Given that these type of users are more familiar with the interaction of multimedia and on-line applications, the design of more interactive tools can highly improved the motivation and success of users searching for children-friendly information.

4.1.4 Cue words analysis

Cue words provide information about the characteristics of the content searched by users and their identification have been proved useful to aid search through query expansion techniques [17]. Our motivation to identify cue words is to identify the most common characteristics and content type of the information searched in the query sets. The identification of cue words is based on a contextual model [33] which is a defined in Equation 3.

$$P(a|w) = \frac{c(a, C(w))}{\sum_i c(i, C(w))} \quad (3)$$

where $C(w)$ is the context of w . This equation models the likelihood of a word a to appear in the context of a given word w . In this paper we define the context of w_i as the set of words that occur at positions $i - 2, i - 1, i + 1, i + 2$. We are interested in mining the context words (cue words) used in the query sets. For this purpose we cluster the words in the queries such that each cluster represents the set of content words w that co-occur with a given context word. Thus, each cluster can be seen as a group of information needs that make use of the same cue word.

The star algorithm is used to perform the clustering by representing the nucleus of each cluster as the context word, the satellites as the content words co-occurring with the context word and the similarity measure between two vertices as the probability given by the model defined in Equation 3. Details about the star-clustering algorithm can be found in [16]. Table 3 shows the top 5 clusters ranked by size for the our three query sets. We found 56, 67 and 69 clusters in the kids, teen and mature teens queries, respectively. Only 5% percent of the cue words of the kids query set appear as cue words in the teens query set, and only 8% in the case of the teens and mature teens query sets. We observed a clear

relation between the clusters and the expected topics of interests of the target users and progression in the topics. For instance social related topics (e.g. life, myspace, boyfriend) are found in big clusters in the teens and mature teens but they are not present in the children's clusters.

4.1.5 Click Analysis

The click information of the datasets was analyzed to compare the retrieval performance between children and general purpose content queries. For this analysis we collected the rank distribution, the mean reciprocal rank (MRR) and the click frequency of the queries in the datasets. Queries in which highly ranked domains are clicked often indicate that information needs could be more efficiently satisfied by the IR system. The MRR of a set of queries Q is defined by Equation 4. Lower MRR values occur when lower ranked documents are more frequently clicked by the user, which indicates poorer retrieval performance than higher MRR values.

$$MRR = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{rank(q_i)} \quad (4)$$

Figure 3 shows the rank frequency distribution of the data sets. This figure demonstrates that the retrieval performance of the queries to retrieve children information is poorer than the queries used to retrieve non-children oriented content, since clicks on lower ranked results are more frequent in the children query set. The MRR values found for the kids, teens, mature teens and whole data set were 0.57, 0.59, 0.58 and 0.73, respectively, which lead to the same conclusions since the MRR value for the whole data set is greater than in the other sets. We also found a drop for the children queries on the clicks ranked below rank 10 in comparison with the whole data set, since only 6% of the queries are below rank 10 for the kids and teens queries (4% for mature teens) while in the whole data 10.3% of the clicks are below 10. This findings suggests that children are less keen to explore low-ranked results beyond the first result page. Although in [13] is reported that children rarely go beyond the top 5 results, we observed a more even distribution in the children clicks rank 10. This findings is also consistent with Bilal's studies [5].

This behavior can be explained by the fact that children tend to explore more results given their lack of focus during the search and their difficulty to specify the information

Table 4: Most frequent queries

Kids	Teens	M_Teens
nickjr.com	the n.com	prom hairstyles
elmo	nasa	american idol
nick jr	hairstyles	cheats
coloring pages	kingdom hearts 2	cheat codes
postopia	claires	prom dresses
candystand	celebrity hairstyles	pussycat dolls
the wiggles	christina aguilera	ea sports
starfall.com	degrassi	bladder infection
dora the explorer	gurl.com	scholarships
primary games	homestarrunner	game cheats

Table 5: Proportion of users across the datasets

	Kids	Teens	M_Teens
Kids	100%	6.0%	9.16%
Teens	13.84%	100%	20.26%
M_Teens	10.7%	10.33%	100%

needed, as is mentioned in [7]. Moreover in [27] is reported that 40% of the children aged 5-15 are willing to explore unseen web results, which also supports the click behavior observed. Statistical significant differences in the MRR and rank mean between the four data sets were found using the Wilcoxon test at 95% confidence level.

4.1.6 Query frequency analysis

Figure 2 shows the cumulative frequency of the queries in our data. We observed a similar behavior in the data sets since queries that appear up to three times in the query log account for 80% of the total number of queries. However, a greater number of unique queries with low frequency are found in the children query sets. This may be due to the fact that children have more problems to represent their information needs causing the formulations of more queries as attempts to fulfill an information need.

Table 4 shows the 10 most frequent queries in the children data³. As it is expected these queries reflected typical interests of the user target groups. For instance, *elmo* and *dora the explorer* are very popular characters among kids. The teens and mature teens queries show a greater interests in video games (*ea sports*), social events (*prom*) and educational matters (*scholarships*) indicating differences in the information needs between kids and older children.

4.1.7 User Analysis

We analyzed the proportion of users that submit queries in more than one of the data sets. The results are summarized in Table 5. This table shows that users that retrieve information for children rarely submit queries to extract information from the teens and mature teens data sets. Analogous results were found for the users of the teens and mature teens datasets. Nonetheless, this result should be taken with caution since the user identification in the AOL query log was performed automatically, which implies that there is no guarantee that users of the query log correspond to actual users.

³Variations of the same domain were removed from this list (e.g nickjr and nickjr.dom).

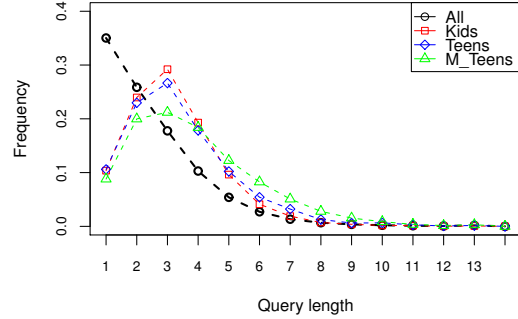


Figure 1: Query length distribution

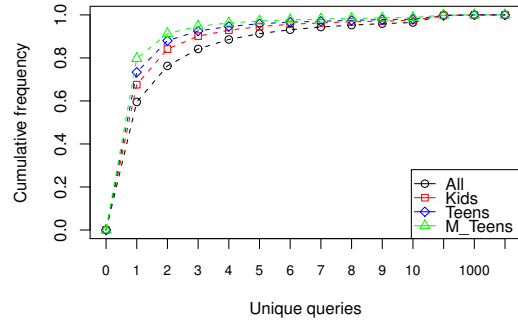


Figure 2: Query frequency distribution

4.2 Session and Goal level results

Sessions allow us to understand the way users accomplish information needs and how they interact with the search engines. We collected three type of metrics to compare the session characteristics of our datasets: session length, duration and query reformulation metrics. Session length is the number of query entries issued in the session. Query entries can refer to new queries issued by the user or to further results clicked using the same query. This metric is an indicator of search efficiency since a greater amount of query entries suggests that more changes to the queries and document visits are needed to fulfill the search task. Session duration is defined as the time in minutes between the last and first query issued in the session and it is an indicator of the complexity of the underlying information need. Query reformulation metrics can be use to understand the way users change their queries to reach their information need. Additionally, goal length and duration were also analyzed as an attempt to understand how efficiently atomic information needs are accomplished and how these goals are issued in our data sets across sessions. Table 6 summarizes the metrics obtained for our data sets. All the results obtained were found statistically significant by using the Wilcoxon test In the following paragraphs these results are analyzed.

4.2.1 Sessions length

Figure 4 shows that general-use sessions are mostly short given that 80% contain less than 5 query entries. On the other hand the length of sessions use to retrieve information for children tend to be longer and its distribution is more

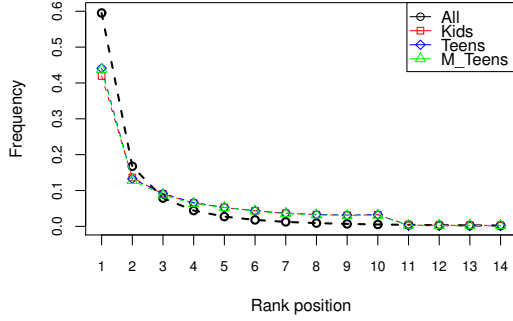


Figure 3: Rank Distribution

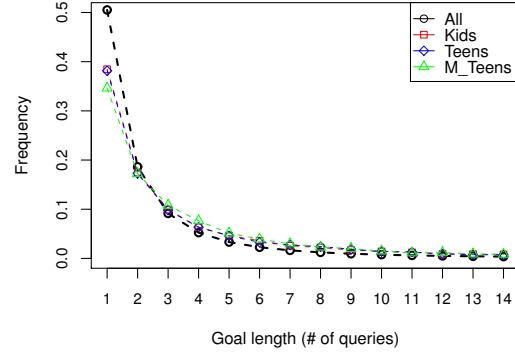


Figure 6: Goals length distribution

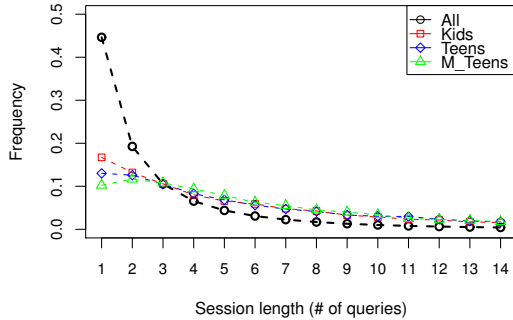


Figure 4: Sessions length distribution

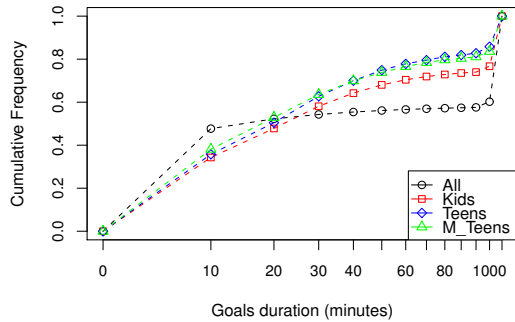


Figure 7: Goals distribution in minutes

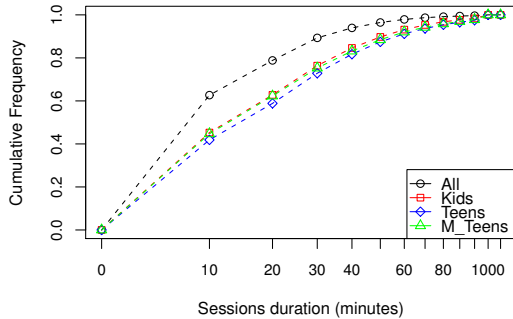


Figure 5: Session duration distribution

Table 6: Summary of sessions (S.) and goals (G.) characteristics

	All	Kids	Teens	M_Teens
S. length (mean)	3.3	8.7	10.1	9.8
S. length (median)	2	5	5	6
S. duration (mean)	11.7	20.4	23.1	21.5
S. duration (median)	5.1	12.5	14.6	12.7
G. length (mean)	4.85	5.79%	5.5	5.8
G. length (median)	1	2	2	2
G. duration (mean)	18,928	5,217	3,763	3,789
G.l duration (median)	13.9	5.2	5.2	4.3

dispersed. The average length found for general-use sessions are in line with previous studies which reported session average of 2.8 [20, 30]. The longer average length found for

the children sessions (see Table 6) suggests that the users of these sessions weren't certain of the relevance of the information found since users have to perform more queries and explore more documents. This result is consistent with Bilal's findings [7] in which children showed *nonlinear navigation style* when solving research tasks. This search style is characterized by the exploration of several choices before a final relevance judgment is made [7]. This result can also indicate that the documents retrieved by the search engine are not sufficient to satisfy the user's information need.

4.2.2 Sessions duration

The duration in minutes of the four types of sessions is shown in Figure 5. This figure shows that users require more time to explore and complete information needs associated to children content than to general-purpose content. Statistically significant differences were found in the average session duration between the kids/teenagers/mature teenagers and general-purpose sessions. No significant differences were found among the types of children sessions. The longer duration of sessions to retrieve information for children suggest more difficulty to solve the information tasks associated to these sessions. This result is in line with the greater amount of pages visited and queries issued found in the session length analysis. This result is also consistent with the trend found in users to click higher ranked pages when children-content queries are issued as it was shown in the previous section. The average session duration found for the children's sessions (20.38 minutes) is in line with the average time reported by Bilal [7] for children that were

Table 7: Frequency of query reformulations

	All	Kids	Teens	M_Teens
n.q	48.1%	25.9%	26.5%	24.5%
w.a	1.5%	1.37%	1.8%	1.8%
w.r	0.1%	0.21%	0.2%	0.2%
w.c	6.3%	7.8%	8%	8.5%
m.r	38.8%	59.6%	57.6%	59%
p.q	3%	3.3%	3.7%	3.5%
s.c	1.8%	1.5%	2.1%	2.1%

unsuccessful completing fact-based information tasks (19.69 minutes). Nonetheless, the long duration found in the children’s sessions can also suggest that broad and open information tasks are more frequent for these users and that in general users are less successful finding information for children. Interestingly, we obtain different conclusions when we calculated the session success rate by considering a *successful search* those sessions that end with a clicked domain, as it is suggested in [9]. We found that the success rate for the kids, teens, mature teens and whole data were 82.5%, 79%, 79.8%, 52.2%, respectively. These values would indicate that users of the children sessions are more successful in their information seeking tasks. However, these values are biased since children tend to explore and click on more results than adults (This behavior will be explored in further detail in the query reformulation analysis). We consider that this metric also evaluates the *trust* children have on Web results, in which more clicks indicate more *trust*. In [27] is reported that only 49% of the children (aged 12-15) make some critical judgment about the truthfulness of the results, which suggests that one of the reasons why children clicked more than adults is their greater belief on the content generated by search engines.

4.2.3 Goals length and duration

The goal length is an indicator of the efficiency to express atomic information needs. On the other hand, since goals were grouped without considering the session limits (without time window restrictions), long goal durations indicate the presence of information needs that involve long term planning, as planning a trip, or information needs that are recurrent, for example checking the news. Figure 6 shows the length distribution of the goals extracted from the data sets. Contrary to the session length behavior, the difference in the length between the datasets is small (4.8 in the whole data set vs 5.8 in the kids set on average). This result indicates that children sessions contain a greater formulation of atomic needs. Figure 7 shows that the goal duration of general-purpose goals are significantly longer than in the children goals. This result is interesting because it suggests that children information needs are less frequently conformed by atomic information needs split in long periods of times. This behavior is consistent with previous experiments in which children are found to be less focus during the search process [6], which diminishes their ability to formulate information needs in longer time periods.

4.2.4 Query reformulation analysis

Users constantly modify their queries in an attempt to get better results from the search engine. The analysis of these query refinements allow us to have a better understanding of the way user’s interact with the search engine

and the search strategies employed to satisfy their information needs. In this paper we analyze the following types of query reformulations:

- Words added to the query (w.a): The previous query is a strict suffix of the target query. e.g. $\{dora\}_{i-1} \xrightarrow{w.a} \{dora\ the\ explorer\}_i$
- Words removed from the query (w.r): Target query is a strict suffix of the previous query. e.g. $\{barbie\ coloring\ pages\}_{i-1} \xrightarrow{w.r} \{barbie\}_i$
- Change of words in the query (w.c): Target query contains at least one word in common with the previous query. Word order is ignored. e.g. $\{all\ about\ spiders\}_{i-1} \xrightarrow{w.c} \{all\ about\ cobras\}_i$
- Spelling correction (s.c): The Levenshtein distance between the target and previous query is ≤ 2 . e.g. $\{candysand\}_{i-1} \xrightarrow{s.c} \{candystand\}_i$
- New query (n.q): Target query does not share any words with the previous query and the Levenshtein distance is greater than 2. e.g. $\{sesame\ street\}_{i-1} \xrightarrow{n.q} \{elmo\}_i$
- More results from the same query (m.r): Target query is identical to previous query and it is used to access a different website.
- Return to previous query (p.q): The target query was submitted during the same session.

Similar query reformulations types have been used in previous query log analysis [28, 19]. Although *m.r* is not a formal query reformulation (since no change is performed on the previous query), we included it in this analysis because this action is commonly use in the search process. Table 7 shows the percentages of the query reformulation types found in our data. These percentages were calculated on sessions containing more than one query which correspond to the 83%, 87%, 90% and 55% of the kids, teens, mature teens and whole data sessions, respectively. A salient difference is the average drop of 22.5% of new queries issued in the children sessions compare to the general-purpose sessions. Most of this drop is reflected in the greater use of the same queries to explore further results, which accounts on average for 90% of the new query reformulation type drop. Small gains in all the other query reformulation types were also found in the teens and mature teens sessions. Nonetheless, kids sessions only showed increase in the word removing, word changing and reusing of previous queries. Although, it has been shown that children commit spelling mistakes more frequently than older users, we didn’t find an increase of the spelling correction reformulation type in the kids sessions. This may be indicate the need to employ more robust methods to detect spelling variations. However, considering that the percentage of spelling corrections is very low in all the session types, this result may be also due to the lack of appropriate spelling corrections tools in the search engine. The rank data between query pairs and the click patterns suggested in [19] were employed to evaluate the effectiveness of the query reformulations. These click patterns are established between a query and its reformulation and can be described in the following way:

Table 8: Click pattern frequency for the query reformulation types

	Dataset	Click-Click	Click-Skip	Skip-Click	Skip-Skip	Mean rank change	Mean time change
n.q	kids	35.9%	22.1%	23.2%	18.6%	-3.8	371.2
	teens	35.7%	22.4%	22%	19.7%	-3.04	359.6
	m.teens	36.7%	22.5%	21.6%	19%	-4.11	371.6
	all	23.7%	19.64%	20.3%	36.2%	-1.98	337.3
w.a	kids	30.6%	14.8%	28.3%	26.1%	-5.83	136.19
	teens	31.6%	13.8%	27.7%	26.8%	-3.34	133.8
	m.teens	30.8%	14.2%	30.7%	24.2%	-7.29	293.8
	all	18.1%	13.2%	28.9%	39.6%	-2.9	117.4
w.r	kids	40%	24.4%	19.5%	16.1%	-5.7	312.1
	teens	48.8%	13.2%	24.1%	13.8%	-3.9	310.5
	m.teens	43.6%	21.8%	21.5%	13%	-4.4	146.6
	all	27.2%	16.5%	24.3%	32%	-3.6	273.7
w.c	kids	42.9%	18.1%	21.3%	17.5%	-6.5	253.6
	teens	39.6%	18.7%	21%	20.5%	-6.1	235.7
	m.teens	40.1%	19.5%	21.9%	17.5%	-5.77	259.8
	all	29.8%	18%	21.5%	30.5%	-3.6	239.3
m.r	kids	77.7%	4.3%	4.7%	13%	2.2	53.2
	teens	73.6%	3.9%	4.5%	17.9%	2.1	49.6
	m.teens	77.7%	4.5%	5%	12.7%	2.31	51.2
	all	59.2%	4.9%	6.4%	29.4%	2.1	69.2
p.q	kids	19.5%	24.6%	24.2%	31.5%	2.2	190.1
	teens	21.4%	23.1%	21.6%	33.8%	0.5	191.5
	m.teens	21.2%	25%	22.5%	31.2%	2	188.3
	all	13.6%	21.1%	18.9%	46.2%	1.3	182
s.c	kids	8.8%	29.9%	32.3%	28.9%	-2.6	52.2
	teens	9.9%	28.6%	32.4%	28.9%	-1.8	48.6
	m.teens	10.3%	31.2%	33.4%	24.9%	-2.38	52.8
	all	5.5%	23.9%	29.1%	41.4%	-1.6	48.6

1. A domain was accessed in the previous query and in the reformulation query (*click_click*)
2. A domain was not accessed in the previous query and in the reformulation query (*skip_click*)
3. A domain was accessed in the previous query but not in the reformulation query (*click_skip*)
4. A domain was accessed in the reformulation query but not in the previous query frequency (*skip_skip*)

Table 8 summarizes the click patterns found for each one of the reformulations types defined above. Patterns (1) and (4) indicate that the reformulation was successful, on the other hand, patterns (2) and (3) indicate that the query reformulation was not effective. The ratio between these patterns also provide important information about the use of the query reformulations and their effectiveness. Low values of the ratio $\frac{(1)+(3)}{(2)+(4)}$ suggests that the user is not satisfied with the initial query and the query reformulation is employed as an attempt to obtain better results. On the other hand, high values of this ratio indicates that users are refining or specializing their queries since a result has already been accessed. The query reformulation types *w.c*, *w.r* and *m.r* have the highest ratio considering all the sessions of the datasets which suggests that users tend to use these types of reformulations to refine their queries, since previous pages were accessed using similar queries. The high ratio value found for *m.r* is logical since users tend to keep the same query when previous results have been shown useful, as it is suggested by the high percentage found in the click-click pattern. Low rate values were found for *w.a*, *p.q* and *s.c* which suggests that these reformulation types are mostly used when users are not satisfied with their current queries.

Similar conclusions were drawn by Huang et al. [19] on this query log by using a more extended set of query reformulations.

The same reformulation types were found to have the lowest ratio values in the children sessions which shows that these reformulation types are also preferred when users are not satisfied with their current queries. Nonetheless, a different behavior was found in the children sessions since all the ratio values are higher than in the general purpose sessions which suggests that query reformulations are more frequently used as follow-up of current results obtained. It is important to mention that the difference in the ratios are mostly caused by the significant higher use of the pattern *click - click* in the children queries. This finding is in-line with [6] studies on children search behavior stating that children need to explore further options before making a final relevance decision. This result also indicates that it is harder for these users to identify the relevancy of the results from the snippet and title presented in the result list, making necessary the exploration of the results in greater detail. This result is also aligned with our previous findings on the children queries rank data, and the children session characteristics since a greater amount of results clicked explains the longer length and duration of children sessions.

The ratio $\frac{(3)+(4)}{(1)+(2)}$ is an indicator of the effectiveness of the query reformulations. The most effective query reformulations were *w.r*, *w.c* and *m.r* and the least effective were *wa*, *pq* and *sc* in the children and general-purpose sessions. However the ratio values in the children sessions were always higher given the greater number of clicks reported in these sessions. The higher rate value of the *p.q* query reformulation in the children sessions can be explained by the fact that children tend to repeat the same search even if it does not return new results [5]. This occurs given their percep-

tion of the web as a source containing all the information they need. This behavior is also in-line with their tendency to loop searches and hyperlinks, as it is reported by Bilal [6].

The mean rank change represents the rank position difference of the clicked domains in the reformulated and original query. Higher rank position values correspond to the results located in the bottom of the result list and lower values to the results on the top of the list. Thus, higher mean rank change indicates that the query reformulation is less successful since users click in average on results below the original queries. We found that all the query reformulations are successful in children and general-purpose sessions, except for *m.r* and *p.q*. This result is logical since these actions do not involve query changes and users generally click on top results first. It is interesting to note that although query reformulation is significantly less used in the children sessions (given the high used of the action *m.r*), these can be highly helpful since the mean rank changed is lower for the children sessions.

The mean time change measures the average time in seconds that users wait to perform a query reformulation. It is measured by calculating the submission time difference in seconds of the reformulated and original query. The wait time to perform query reformulations is longer for most types in the children sessions. This results was expected since children’s physical and cognitive skills are less developed than in grown ups. (e.g. type speed and reading skills are lower), moreover children need more time to concrete their information needs and they are less focus during the search [7]. Surprisingly the only reformulation type that doesn’t follow this trend is the *m.r* type, which is the most frequently used in the children sessions. However, we found in the query log cases in which consecutive identical queries are submitted with the same time-stamp, which influences the accuracy of this measure. We found that this occurs in 10% of the *m.r* actions in the sessions of the entire query log and 30% in *m.r* actions in the kids sessions.

5. LESSONS LEARNED

The better understanding of users retrieving information for children on a large-scale achieved in this paper allow us to discuss several ways to improve the search experience of these users. We discuss improvements on two IR dimensions: query assistance and aggregated search.

5.1 The need for query assistance

The use of longer queries to retrieve children-friendly content can be one of the causes of the lower retrieval performance found for these queries since it has been shown that longer queries have poorer performance in current search engines [23]. For this reason we consider that refining long queries is highly beneficial for these users. Kumeran et al.[4] present a method to rewrite shorter queries by segmenting the original query and ranking the segments. Their method obtained 8% MAP improvement on two TREC collections (TREC123 and Robust 2004). Similarly, Bendersky et al. [3] employ a method to extract key concepts from the queries based on a classifier trained with query-dependant and corpus features. These keywords are used to rewrite shorter queries which are also proven to have better retrieval performance on a TREC collection.

The greater number of adjectival and verbal phrases found

in the children queries suggests that the creation of query formulation tools based on phrases instead of keywords can also be beneficial for our target users as Jensen et al. [21] shows for the general Web user. Additionally, query rewriting methods may also be explored to rewrite phrases to keywords or entities to boost the retrieval performance

Cue words are also a valuable resource to rewrite queries by using the cues terms associated to the contexts words or to the entities that occur in the query. This method can ease the exploration of information by providing different content types and dimensions of the topic being searched. The study of further query rewriting techniques as the one presented in [34] can also be beneficial to reduce the cognitive load of reformulating queries which are rarely used in the kids, teens and mature teens data sets. Finally, query assistance functionality should also help and train children to improve their skills to search the web.

5.2 Difficulty of searching for kids

The low percentage of informational queries found in the *children queries* suggests that although these users are familiar with interactive applications, they are not fully harvesting the information content available in the Web. This behavior may be due to the lack of expertise formulating information needs using keywords, to the difficulty of identifying relevant information from the web results or to the lack of more friendly methods to guide the search. This difficulty was also observed in the greater number of entries and longer duration of sessions. This finding may also be due to the preference of these users to use the internet for entertainment purposes as it is observed in the clusters presented in Table 3 and as it is reported in [27]. This user search behavior suggests that more efficient ways to gather the information and more efficient ways to present it to the user are required for these type of users. We consider that aggregated search represents a promising paradigm to address these difficulties. Aggregated search refers to the selection of results from diverse sources and content types and the integration of this information to aid the user to reach his/her information need more efficiently [26].

The identification and clustering of cue words shown in this paper has potential to assist the selection of relevant verticals for the information needs of children. Although current IR systems for children as *Yahoo Kids!* or *Ask Kids* already offer categories associated to some of the cue words we identified in the clusters (e.g. games, coloring pages, poems, jokes, homework help, etc.), the dynamic selection of verticals for each query is still very limited. The detection of cue words from the submitted query and from historical data (query logs) also provide hints to the type and characteristics of the information to be displayed. This information is useful to complement vertical-selection classifiers as the one presented in [1] by designing methods to match these cue words to the available verticals.

Additionally, cue words are well-suited to be used in the information integration phase of the aggregated search paradigm. In this phase content from different verticals is processed, reorganized and presented to the user. Current IR systems only provide very simple methods to aggregate results from the verticals. IR systems could parse the content of the web results and aggregate only the content type suggested by the cue words of the query, this would highly reduce the cognitive load of users to find relevant information.

6. ACKNOWLEDGEMENTS

This research is funded by the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. 231507

7. CONCLUSIONS AND FUTURE WORK

In this study we found differences with statistical significance in queries and sessions employed to retrieve information for children and general-purpose content. Given that most of our findings are in-line with previous studies of children's information-seeking behavior [5, 6, 7, 13], we consider that this work represents a valuable and well-suited methodology to study the search behavior of users retrieving information for children. Although these case-studies are highly valuable for the understanding of user's behavior, the characterization of users on a large-scale using query logs has several advantages: It is unobtrusive (which makes possible to capture the actual user behavior), safer, repeatable, non-disruptive, non-reactive, inexpensive, a resource of longitudinal data (accessibility over long periods of time) and it allows the design and evaluation of solutions in realistic scenarios. Although the main disadvantage of query logs is the presence of noise, this problem can be overcome by using accumulated results from samples of reasonable size.

Besides corroborating previous case-studies results on a large scale, we were also able to characterize the tasks children prefer in the Web (by the classification of query intent), the preferred topics/interest of these users (by the cue words analysis) and their query reformulation behavior.

It is important to mention that the design of search success measures on query logs need to be addressed since current metrics based on clicks (e.g. mean reciprocal rank, click pattern analysis) are biased given the greater click action of children, which can be an indicator of *trust* on Web content instead of search success. As it is pointed out in [7], search success should go beyond information access rate. As future work we will study query assistance and content integration methods to improve the search success and experience of children by using the information identified in this paper (e.g. query classification query structure, query intent, query reformulations and session activity). Additionally, we plan to apply this study in more recent query logs to corroborate the results we have obtained.

8. REFERENCES

- [1] J. Arguello, F. Diaz, J. Callan, and J.-F. Crespo. Sources of evidence for vertical selection. In *SIGIR '09*, pages 315–322, New York, NY, USA, 2009. ACM.
- [2] R. Baeza-Yates, L. Calderón-Benavides, and C. González-Caro. The intention behind web queries. pages 98–109. 2006.
- [3] M. Bendersky and W. B. Croft. Discovering key concepts in verbose queries. In *SIGIR '08*, pages 491–498, New York, NY, USA, 2008. ACM.
- [4] M. Bendersky and W. B. Croft. Analysis of long queries in a large scale search log. In *WSCD '09*, pages 8–14, New York, NY, USA, 2009. ACM.
- [5] D. Bilal. Children's use of the yahooligans! web search engine: I. cognitive, physical, and affective behaviors on fact-based search tasks. *JASIS*, 51(7):646–665, 2000.
- [6] D. Bilal. Children's use of the yahooligans! web search engine II. cognitive and physical behaviors on research tasks. *J. Am. Soc. Inf. Sci. Technol.*, 52(2):118–136, 2001.
- [7] D. Bilal. Children's use of the yahooligans! web search engine III. cognitive and physical behaviors on fully self-generated search tasks. *J. Am. Soc. Inf. Sci. Technol.*, 53(13):1170–1183, 2002.
- [8] P. Boldi, F. Bonchi, C. Castillo, D. Donato, A. Gionis, and S. Vigna. The query-flow graph: model and applications. In *CIKM '08*, pages 609–618, New York, NY, USA, 2008. ACM.
- [9] D. J. Brenes and D. Gayo-Avello. Stratified analysis of aol query log. *Inf. Sci.*, 179(12):1844–1858, 2009.
- [10] A. Broder. A taxonomy of web search. *SIGIR Forum*, 36(2):3–10, 2002.
- [11] K. C. C. Inside the search process: Information seeking from the user's perspective, 1991.
- [12] S. Chien and N. Immorlica. Semantic similarity between search engine queries using temporal correlation. In *WWW '05*, pages 2–11, New York, NY, USA, 2005. ACM.
- [13] A. Druin, E. Foss, L. Hatley, E. Golub, M. L. Guha, J. Fails, and H. Hutchinson. How children search the internet with keyword interfaces. In *IDC '09*, pages 89–96, New York, NY, USA, 2009. ACM.
- [14] V. Ermolayev, N. Keberle, S. Plaksin, and V. Vladimirov. Capturing semantics from search phrases: Incremental user personification and ontology-driven query transformation. In *ISTA*, pages 9–20, 2003.
- [15] C. for Public Broadcasting. Connected to the future: A report on children's internet use from the corporation for public broadcasting, 2002.
- [16] R. Gil-García and A. Pons-Porrata. Hierarchical star clustering algorithm for dynamic document collections. In *CIARP '08*, pages 187–194, Berlin, Heidelberg, 2008. Springer-Verlag.
- [17] S. Guo and N. Ramakrishnan. Mining linguistic cues for query expansion: applications to drug interaction search. In *CIKM '09*, pages 335–344, New York, NY, USA, 2009. ACM.
- [18] D. He and A. Göker. Detecting session boundaries from web user logs. In *In Proceedings of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research*, pages 57–66, 2000.
- [19] J. Huang and E. N. Efthimiadis. Analyzing and evaluating query reformulation strategies in web search logs. In *CIKM '09*, pages 77–86, New York, NY, USA, 2009. ACM.
- [20] B. J. Jansen, D. L. Booth, and A. Spink. Determining the informational, navigational, and transactional intent of web queries. *Inf. Process. Manage.*, 44(3):1251–1266, 2008.
- [21] E. C. Jensen, S. M. Beitzel, A. Chowdhury, and O. Frieder. Query phrase suggestion from topically tagged session logs. In *FQAS*, pages 185–196, 2006.
- [22] R. Jones and K. L. Klinkner. Beyond the session timeout: automatic hierarchical segmentation of search topics in query logs. In *CIKM '08*, pages 699–708, New York, NY, USA, 2008. ACM.
- [23] G. Kumaran and V. R. Carvalho. Reducing long queries using query quality predictors. In *SIGIR '09*, pages 564–571, New York, NY, USA, 2009. ACM.
- [24] T. Z. J. M. Barbaro. A Face is exposed for AOL searcher no. 4417749. The New York Times, 2006. <<http://www.nytimes/2006/08/09/technology/09aol.html>>.
- [25] S. L. M. Bober. Uk children go online : surveying the experiences of young people and their parents. London:LSE Research Online, 2004.
- [26] V. Murdock and M. Lalmas. Workshop on aggregated search. *SIGIR Forum*, 42(2):80–83, 2008.
- [27] Ofcom. Uk children's media literacy: Research document, March 2010.
- [28] G. Pass, A. Chowdhury, and C. Torgeson. A picture of search. In *InfoScale '06: Proceedings of the 1st international conference on Scalable information systems*, page 1, New York, NY, USA, 2006. ACM.
- [29] S. R.S. *Children's thinking 2d ed.* Englewood Cliffs. Prentice Hall, 1991.
- [30] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz. Analysis of a very large web search engine query log. *SIGIR Forum*, 33(1):6–12, 1999.
- [31] A. Spink, D. Wolfram, M. B. J. Jansen, and T. Saracevic. Searching the web: the public and their queries. *J. Am. Soc. Inf. Sci. Technol.*, 52(3):226–234, 2001.
- [32] V. A. Walter. The information needs of children. *Advances in Librarianship*, pages 111–129, 1994.
- [33] X. Wang and C. Zhai. Mining term association patterns from search logs for effective query reformulation. In *CIKM '08*, pages 479–488, New York, NY, USA, 2008. ACM.
- [34] W. V. Zhang, X. He, B. Rey, and R. Jones. Query rewriting using active learning for sponsored search. In *SIGIR '07*, pages 853–854, New York, NY, USA, 2007. ACM.