Hourly Analysis of a Very Large Topically Categorized Web Query Log

Steven M. Beitzel, Eric C. Jensen, Abdur Chowdhury, David Grossman, Ophir Frieder
Illinois Institute of Technology
10 W 31 St.
Chicago, IL 60616
{steve,ej,abdur,grossman,frieder}@ir.iit.edu

ABSTRACT

We review a query log of hundreds of millions of queries that constitute the total query traffic for an entire week of a general-purpose commercial web search service. Previously, query logs have been studied from a single, cumulative view. In contrast, our analysis shows changes in popularity and uniqueness of topically categorized queries across the hours of the day. We examine query traffic on an hourly basis by matching it against lists of queries that have been topically pre-categorized by human editors. This represents 13% of the query traffic. We show that query traffic from particular topical categories differs both from the query stream as a whole and from other categories. This analysis provides valuable insight for improving retrieval effectiveness and efficiency. It is also relevant to the development of enhanced query disambiguation, routing, and caching algorithms.

Categories and Subject Descriptors: H.3.5 [**Information Storage and Retrieval**]: Online Information Services – *Web-based services*

General Terms: Measurement, Human Factors.

Keywords: Query Log Analysis, Web Search.

1. INTRODUCTION

Understanding how queries change over time is critical to developing effective, efficient search services. We are unaware of any log analysis that studies differences in the query stream over the hours in a day; much less how those differences are manifested within topical categories. We focus on Circadian changes in popularity and uniqueness of topical categories. Emphasis on changing query stream characteristics over this longitudinal (time) aspect of query logs distinguishes this work from prior static log analysis, surveyed in [7].

We began with the hypothesis that there are very different characteristics during peak hours and off-peak hours during a day. After reviewing a week's worth of data – hundreds of millions of queries - we have found, not surprisingly, that:

SIGIR'04, July 25–29, 2004, Sheffield, South Yorkshire, UK.

- *The number of queries issued is substantially lower during non-peak hours than peak hours.*

However, we knew little about how often queries are repeated from one hour of the day to the next. After examining the behavior of millions of queries from one hour of the day to the next we have found the less obvious result:

- *The average number of query repetitions in an hour does not change significantly on an hourly basis throughout the day.*
- *Most queries appear no more than several times per hour.* These queries consistently account for a large portion of total query volume throughout the course of the day.
- *The queries received during peak hours are more similar to each other than their non-peak hour counterparts.*

We also analyze the queries representing different topics using a topical categorization of our query stream. These cover approximately 13% of the total query volume. We hypothesized that traffic behavior for some categories would change over time and that others would remain stable. For 16 different categories, we examined their traffic characteristics:

- *Some topical categories vary substantially more in popularity than others as we move through an average day.* Some topics are more popular during particular times of the day, while others have a more constant level of interest over time.
- *The query sets for different categories have differing similarity over time.* The level of similarity between the actual query sets received within topical categories varies differently according to category.

This leads us to believe that predictive algorithms that are able to estimate the likelihood of a query being repeated may well be possible. This could have a significant impact on future cache management and load-balancing algorithms. Such algorithms could improve retrieval effectiveness by assisting in query disambiguation, making it easier to determine what information need is being expressed by a query at a given time. They could also assist research in search efficiency that takes into account query arrival-rates [3].

Our analysis covers the entirety of the tens of millions of queries each day in the search log from America Online™ over a complete week in December. This represents a population of tens of millions of users searching for a wide variety of topics. Section 2 reviews the prior work in query log analysis. Section 3 describes our analysis of overall query traffic. Section 4 describes our analysis of trends in categorized queries. Finally, in Section 5 we present our conclusions and directions for future work.

2. PRIOR WORK

Examinations of search engine evaluation indicate that performance likely varies over time due to differences in query sets and collections [6]. Although the change in collections over time has been studied (e.g., the growth of the web) [10], analysis of users' queries has been primarily limited to the investigation of a small set of available query logs that provide a snapshot of their query stream over a fixed period of time. Prior work can be partitioned into static query log analysis and some recent disclosures by web search engines.

Query log analysis can be partitioned into large-scale log analysis, small-scale log analysis and some other applications of log analysis such as categorization and query clustering. Jansen and Pooch provide a framework for static log analysis, but do not address analysis of changes in a query stream over time [7]. Given that most search engines receive on the order of between tens and hundreds of millions of queries a day [22], current and future log analysis efforts should use increasingly larger query sets to ensure that prior assumptions still hold.

Previous studies measured overall aspects of users' queries from static web query logs. In the only large-scale study (all others involve only a few million queries), Silverstein concludes that users typically view only the top ten search results and that they generally enter short queries from a static analysis of an AltaVista query log from six weeks in 1998 consisting of 575 million non-empty queries [16]. He also found that only 13.6% of queries appear more than three times, the top 25 queries represent 1.5% of the total query volume, and in 75% of sessions users do not revise their queries. Additionally, co-occurrence analysis of the most frequent 10,000 queries showed that the most correlated terms are often constituents of phrases. No time-based or topic-based analysis of this query load was reported; it does not provide insight into how or when any usage or topical interest changes occur. Other studies examine the effect of advanced query operators on the search service coverage of Google, MSN, and AOL, finding that in general, they had little effect [4]. These overall statistics do not provide any insight into temporal changes in the query log, but do provide some insight into how people use search services.

Jansen, et. al, also provide analysis of query frequency [7][19]. Their findings indicate that the majority (57%) of query terms from the Excite log of more than 51,000 queries are used only once, and a large majority (78%) occur three times or less. These studies show that neither queries nor their component terms follow a Zipfian distribution, as the number of rare, infrequently repeated queries and terms is disproportionately large. Other studies have focused on user behavior at the query session level and found varying results, with some estimating reformulated queries constituting 40-52% of queries in a log [18][21]. Wang, et. al examined a log of more than 500,000 queries to a university search engine from 1997-2001 [23]. They find trends in the number of queries received by season, month, and day. We extend upon this work by examining the larger community of general web searchers and analyzing trends corresponding to hour of day.

Several studies examine query categories in small, static logs. Spink, et. al analyzed logs totaling more than one million queries submitted to the Excite web search engine during single days in 1997, 1999, and 2001 [18][19][20]. They classified approximately 2,500 queries from each log into 11 topical categories and found that although search topics have changed over the years, users' behaviors have not. Ross and Wolfram categorized the top 1,000 term pairs from the one million query Excite log into 30 subject areas to show commonalities of terms in categories [13]. Jansen, et. al used lists of terms to identify image, audio, and video queries and measure their presence in the one million query Excite log [9]. In order to examine the differences in queries from users in different countries, Spink, et. al, examined a 500,000 query log from the FAST web search engine during 2001, believed to be used largely by Europeans at that time, classifying 2,500 queries from it into the same topical categories. They found differences between FAST and Excite in the topics searched for [17].

Other work manually grouped queries by task. Broder defines queries as informational, navigational or transactional and presents a study of AltaVista users via a popup survey and manual categorization of 200 queries from a log [2]. Beitzel, et. al implicitly categorized queries from a search log as navigational by matching them to edited titles in web directories to automatically evaluate navigational web search [1]. Xie and Wolfram automatically categorized query terms by using results from web search engines to assign the terms to broad subject categories [25].

Several studies of query caching examine query frequency distributions from a static log, focusing on the average likelihood of an arbitrary query being repeated over the entire, fixed-length log. Lempel and Moran evaluated the performance of caching strategies over a log of seven million queries to AltaVista in 2001 and found that the frequencies of queries in their log followed a power law [11]. Eiron and McCurley compared query vocabulary from a log of nearly 1.3 million queries posed to a corporate intranet to the vocabulary of web page anchor text and found that the frequency of queries and query terms follows a tail-heavy power law [5]. Xie and O'Hallaron studied query logs from the Vivisimo meta-search engine of 110,881 queries over one month in 2001 in comparison to the Excite log of 1.9 million over one day in 1999 and found that although as in other studies over half of the queries are never repeated, the frequencies of queries that are repeated do follow a Zipfian distribution [26]. Saraiva, et. al evaluated a two-level caching scheme on a log of over 100,000 queries to a Brazilian search engine and found that query frequencies follow a Zipf-like distribution [15]. Markatos simulated the effect of several types of query caches on an Excite query log of approximately one million queries and found that traditional caching methods provide significant improvements in efficiency [12]. Although traditional MRU-style caches obviously enhance throughput by exploiting temporal locality at the minute-to-minute level, these studies do not examine changes in the query stream according to the hour of the day that may be leveraged in enhanced cache design.

It is well known that different users represent the same information need with different query terms, making query clustering attractive when examining groups of related queries. However, as Raghavan and Sever have shown, traditional similarity measures are unsuitable for finding query-to-query similarity [13]. Wen, et. al, incorporated click-through to cluster users' queries [23]. In evaluating their system, they analyzed a random subset of 20,000 queries from a single month of their approximately 1-million queries-per-week traffic. They found that the most popular 22.5% queries represent only 400 clusters of queries using differing sets of query terms.

Many web search services have begun to offer views of the most popular and/or changing (becoming drastically more or less popular) queries: AOL – *Member Trends*, Yahoo - *Buzz Index*, Lycos - *The Lycos 50 with Aaron Schatz*, Google – *Zeitgeist*, AltaVista - *Top Queries*, Ask Jeeves, Fast (AllTheWeb). These views necessarily incorporate a temporal aspect, often showing popular queries for the current time period and those that are consistently popular. Some also break down popularity by topical categories. Systems seeking to display changing queries must address the issue of relative versus absolute change in a query's frequency to find queries whose change is "interesting", not simply a query that went from frequency one to two (a 200% jump), or one that went from 10,000 to 11,000 (a 1000 absolute change).

3. OVERALL QUERY TRAFFIC

We examine a search log consisting of hundreds of millions of queries from a major commercial search service over the seven-day period from 12/26/03 through 1/1/04. This log represents queries from approximately 50 million users. We preprocess the queries to normalize the query strings by removing any case differences, replacing any punctuation with white space (stripping advanced search operators from the approximately 2% of queries containing them), and compressing white space to single spaces. The average query length is 1.7 terms for popular queries and 2.2 terms over all queries. On average, users view only one page of results 81% of the time, two pages 18% and three or more 1% of the time. First, we examine trends in the query stream as a whole, and then focus on trends related to queries manually categorized into topical categories.

We begin our analysis of the overall stream by examining how the volume of query traffic changes as we move from peak to non-peak hours. We show the percentage of the day's total and distinct number of queries for each hour in the day on average over our seven-day period in Figure 1 (all times in our query log are Eastern Standard Time). Only 0.75% of the day's total queries appear from 5-6AM, whereas 6.7% of the day's queries appear from 9-10PM. Perhaps more interestingly, the ratio of distinct to total queries in a given hour is nearly constant throughout the day. This shows that the average number of times a query is repeated is virtually constant over the hours in a day, remaining near 2.14 with only a 0.12 standard deviation.

Although the average repetition of queries remains nearly constant, we can examine this in greater detail by measuring the frequency distribution of queries at various hours in the day, as seen in Figure 2. From this analysis it is clear that the vast majority of queries in an hour appear only one to five times and that these rare queries consistently account for large portions of the total query volume throughout the course of the day.

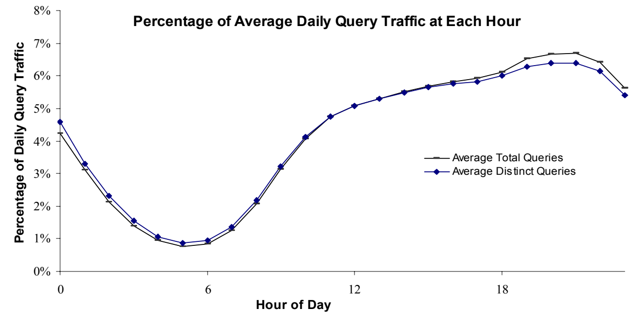

Figure 1

Although we have shown that the query distribution does not change substantially over the course of a day, this does not provide insight into how the sets of queries vary from one hour to the next. To examine this, we measure the overlap between the sets of queries entered during those hours. We use traditional set and bag overlap measures as given in Equation 1 and Equation 2, respectively. Distinct overlap measures the similarity between the sets of unique queries from each hour, while overall (bag) overlap measures the similarity of their frequency distributions by incorporating the number of times each query appears in an hour, $C(q_i;A)$. While these measures examine the similarity of the sets of queries received in an hour and the number of times they are entered, they do not incorporate the relative popularity or ranking of queries within the query sets. To examine this, we also measure the Pearson correlation of the queries' frequencies. As can be seen from Equation 3 (where $\overline{C(q;A)}$ is the mean number of query repetitions in period A and $s_{C(q;A)}$ is the standard deviation of all the query frequencies in period A), this measures the degree of linear correlation between the frequencies of the queries in each hour, so two hours that had exactly the same queries with exactly the same frequencies would have a correlation of one. Note that this normalizes for the effect of differing query volume, i.e., the correlation of two hours with exactly the same underlying query distributions simply scaled by a constant would also have a correlation of one.

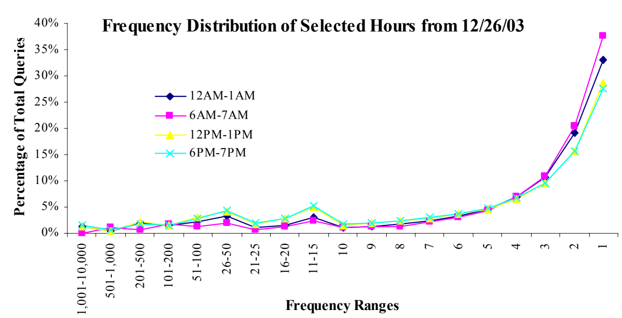

Figure 2

$$dist.overlap\ (A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Equation 1: Distinct Overlap of Query Sets from Hours A and B

$$overlap(A,B) = \frac{\sum_{q_i \in A \cup B} \min(C(q_i;A), C(q_i;B))}{\sum_{q_i \in A} C(q_i;A) + \sum_{q_i \in B} C(q_i;B) - \sum_{q_i \in A \cup B} \min(C(q_i;A), C(q_i;B))}$$

Equation 2: Overall Overlap of Query Sets from Hours A and B

$$r_{A,B} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(C(q_i;A) - \overline{C(q;A)})(C(q_i;B) - \overline{C(q;B)})}{s_{C(q;A)}s_{C(q;B)}}$$

Equation 3: Pearson Correlation of Query Frequencies from Hours A and B

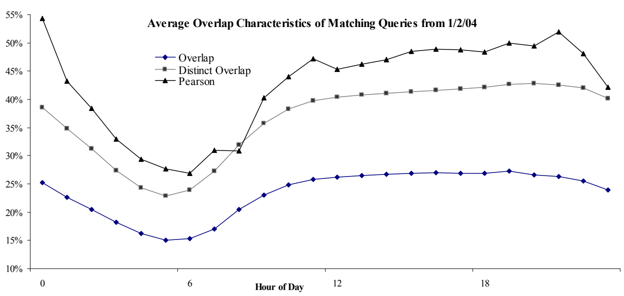

Figure 3

In Figure 3 we examine the average level of overlap and correlation between the query sets received during the same hour for each day over our week. As measuring overlap over the set of all queries appearing in our week would be computationally expensive, we use the set of all the tens of millions of queries in the day after our seven-day period as an independent sample and measure overlap at each hour in our week of the queries matching those in that sample. Although we previously saw that the frequency distribution of queries does not substantially change across hours of the day, Figure 3 shows that the similarity between the actual queries that are received during each hour does in fact change. This trend seems to follow query volume, which is apparent if we sort the same overlap data by query volume as is done in Figure 4. Clearly, as query volume increases the queries that compose that traffic are more likely to be similar across samples of those peak time periods.

This finding is consistent with prior analyses of web query caches showing they significantly improve performance under heavy load. The more redundancy they are able to detect, the more caching algorithms are able to enhance throughput. Although the prior work primarily measures the effect of this redundancy in cache performance, it is obvious that redundancy must exist and be detected for caching to succeed. By examining the overall query stream by hour we are able to infer the effectiveness of general caching algorithms at those times.

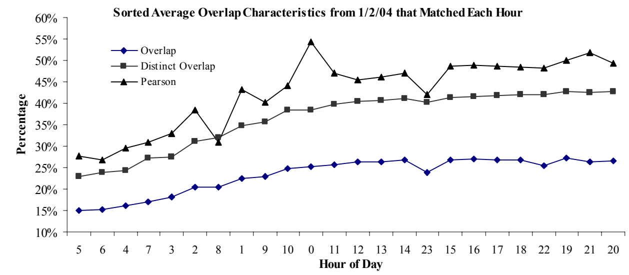

Figure 4

4. QUERY CATEGORIES

In Section 3 we analyzed the entire query log. However, this blanket view of the query traffic does not provide insight into the characteristics of particular categories of queries that might be exploited for enhanced efficiency or effectiveness. For example, a search provider who returns specialized results for entertainment queries cannot determine from general query traffic alone whether a given query is more likely to be referring to entertainment related content or how to best process and cache that query.

The remainder of our analysis focuses on trends relating to topical category of queries. Our query set is categorized simply by exactly matching queries to one of the lists corresponding to each category. These lists are manually constructed by editors who categorize real users' queries, generate likely queries, and import lists of phrases likely to be queries in a category (e.g., cities in the US for the US Sites category). Queries that match at least one category list comprise 13% of the total query traffic on average. This represents millions of queries per day.

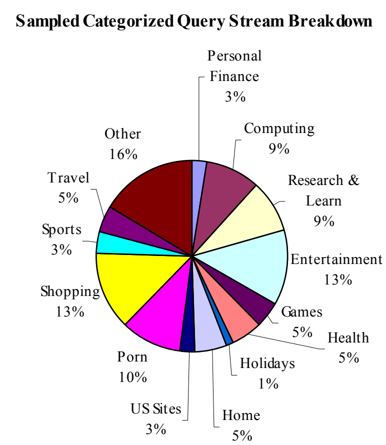

Figure 5

To verify that our defined category lists sufficiently cover the topics in the query stream, we manually classified a random sample of queries, assigning them to "Other" if they did not intuitively fit into an existing category, as can be seen in Figure 5. To determine the number of queries required to achieve a representative sample, we calculate the necessary sample size in queries, $ss = (z^2\sigma^2)/\beta^2$, where z is the confidence level value, σ is the sample standard deviation, and β is the error rate. By setting our confidence level to 99% and error rate to 5%, we require a sample of 600 queries. The relative percentages for each category of the approximately 13% of query volume that match any category list over our week (see Figure 9) are within the error rate of those from our manually categorized sample. This shows that our lists are a reasonable representation of these topical categories.

We focus on a subset of these categories and examine music and movies independent of other entertainment queries. The relative size of each category list we used is given in Figure 6. Obviously, not all queries listed actually match those entered by users, especially when the category contains large imported lists of phrases.

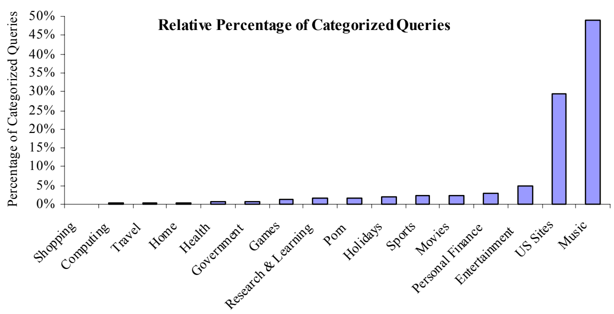

Figure 6

Although we have shown that our lists are a fair representation of the topics in the query stream, this does not indicate what portion of the frequency distribution of that stream they represent. To determine this, we measured the average proportion of queries matching any category list that appear at various frequencies each hour and compared them to the average overall hourly frequency distribution of the query stream (see Figure 7). Unsurprisingly, this comparison shows that queries in the category lists represent more popular, repeated queries than average, although the general shape of the distributions is similar.

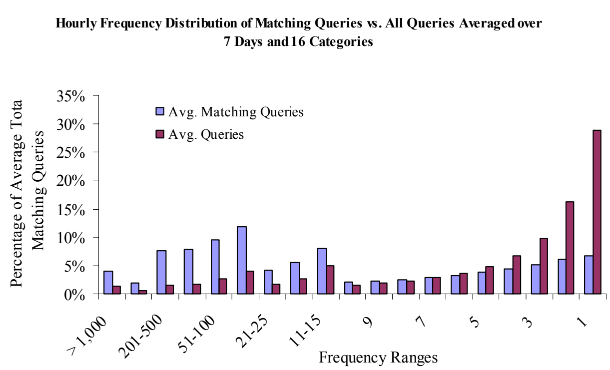

Figure 7

4.1 Trends in Category Popularity

We begin our temporal analysis of topical categories by measuring their relative popularity over the hours in a day. First, we examine the percent of total query volume matching a selected group of category lists, as can be seen in Figure 8. It is clear that different topical categories are more and less popular at different times of the day. Personal finance, for example, becomes more popular from 7-10AM, while music queries become less popular. Although it is difficult to compare the relative level of popularity shift from one category to another due to the differences in scale of each of their percentages of the query stream, it is clear that some categories' popularity changes more drastically throughout the day than others.

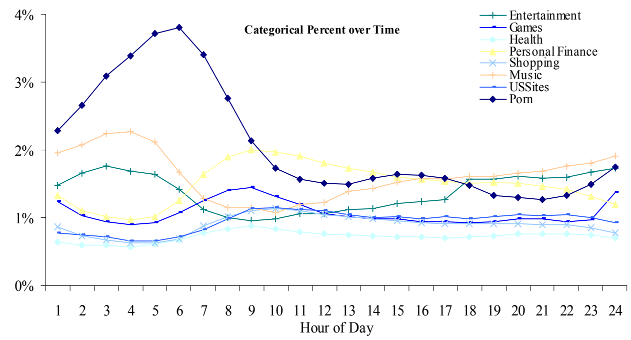

Figure 8

In order to quantify this, we calculated the KL-divergence (Equation 4) between the likelihood of receiving any query at a particular time and the likelihood of receiving a query in a particular category, as can be seen in Figure 9. This reveals that the top three categories in terms of popularity are pornography, entertainment, and music.

$$D(p(q \mid t) \| p(q \mid c,t)) = \sum_{q} p(q \mid t) \log \frac{p(q \mid t)}{p(q \mid c,t)}$$

Equation 4: KL-Divergence of Query Occurrence Likelihood for Category c and Total Stream at Time t

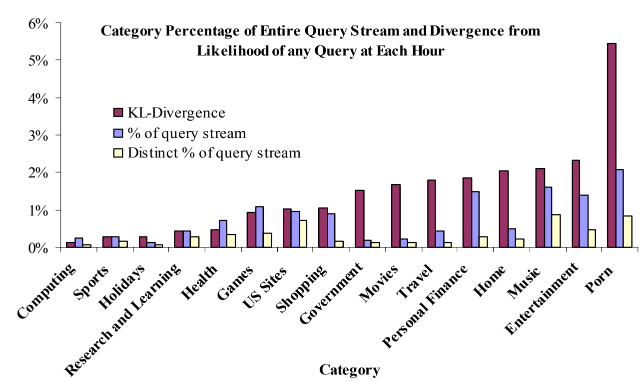

Figure 9

Comparing these divergences to the proportion of categorized queries in each category in Figure 6 quickly illustrates that divergence is not correlated with the number of queries categorized in each category. Also shown in Figure 9 is the average percentage of the entire query volume and distinct queries that match each category. Although the categories that cover the largest portions of the query stream also have the most relative popularity fluctuation, this correlation does not continue throughout all categories.

We drilled down into the highly fluctuating categories and examined the behavior of the queries with the most highly fluctuating frequencies in each category. From this we hoped to gain some insight into the reasons why certain categories fluctuate, and the effect of terms and queries with very high flux on those categories. For example, the three most changing queries for the entertainment category on average over our week were:

Table 1: Top Three Fluctuating Entertainment Queries

gwyneth paltrow
paris hilton
orlando bloom

All three of these queries are specifically related to recent events in US popular culture; the actress Gwyneth Paltrow recently married in secret, and the news of her nuptials broke during the week we analyzed. Hilton Hotel heiress Paris Hilton has been a popular topic recently; she starred in a prime time reality TV show entitled "*The Simple Life*". Also popular is Orlando Bloom, the actor who portrays a popular character in the "*Lord of the Rings"* trilogy. As the final installment of the series was released in US theatres during the week prior to our query log, it is no surprise to see his name as a top-changing query.

Drilling down further, we pinpointed some of the specific instances where these popular queries jumped the most. For example, in the afternoon of Friday, December 27th, the popularity of the query "gwyneth paltrow" skyrocketed. From 3-4PM, it occurred once, from 4-5PM it occurred 67 times, and from 5PM-6PM it occurred 11,855 times. The top changing (on average) twenty-five queries, after normalization, in the Entertainment and Music categories are shown in Table 2.

Table 2: Top 25 Fluctuating Queries from Music and Entertainment

Music	Entertainment
lyrics	gwyneth paltrow
music	paris hilton
britney spears	orlando bloom
furniture	espn
love	disney
hilary duff	johnny depp
good charlotte	much music
sloppy seconds	disney channel
jessica simpson	hgtv
b2k	disneychannel com
eminem	www disneychanel com
christina aguilera	katie holmes pictures
simple plan	pamela anderson
justin timberlake	cartoon network
free music	hilary duff
linkin park	fake
michael jackson	chad michael murray
beyonce	vivica a fox
jennifer lopez	disneychannel
50 cent	care bears
kinky	sailor moon
napster	www cartoonnetwork com
chic	days of our lives
tupac	charmed
blink 182	tom welling

We also looked at some of the most frequently changing terms to see how they relate to the change of entire queries containing those terms. Some excellent examples of this behavior in the Entertainment category include the terms "pictures" (the tenth-most changing term) and "duff" (the 17th-most changing term). We looked at the popularity change (i.e., change in frequency) for queries containing these terms and found that several of them also exhibited large changes over time. For example, on the afternoon of December 28th from noon to 5PM EST, the query "hilary duff" changed from an initial frequency of 27 from 12-1PM to a peak of 131 (from 3-4PM), and then stabilized around 70 for the rest of the evening; similar spikes in frequency for this query occurred at similar times during other days in our period of study.

4.2 Trends in Uniqueness of Queries Within Categories

Although we have shown that different categories have differing trends of popularity over the hours of a day, this does not provide insight into how the sets of queries within those categories change throughout the day. In order to examine this, we return to the overlap measures used in Section 3. Overlap, distinct overlap, and the Pearson correlation of query frequencies for Personal Finance and Music are shown in Figure 10 and Figure 11.

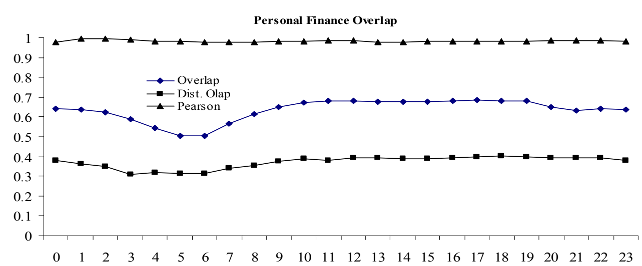

Figure 10

Although the uniqueness of queries in categories in general appears to be correlated with that of the entire query stream (Figure 3), that of particular categories appears to be substantially different from one to the next. For example, if we compare the overlap characteristics of personal finance with those of music, we see they are quite different. Not only does personal finance have generally higher overlap, but it has a much higher overall overlap than distinct overlap, whereas they are nearly equal for music. Other categories with generally high overlap and distinct overlap are shopping, computing, and travel. Also, the correlation of frequencies of personal finance queries is very high all day, indicating searchers are entering the same queries roughly the same relative amount of times, this is clearly not true for music. Some categories have a high Pearson correlation. This indicates that a significant portion of the queries in these categories is often ranked similarly by frequency. These categories are: pornography, travel, research and learning, and computing, and their Pearson correlations are illustrated in Figure 12.

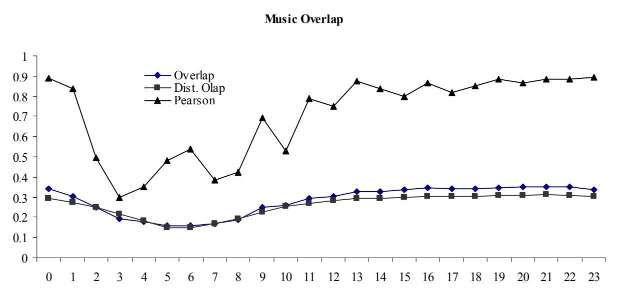

Figure 11

It is clear that some categories have very similarly ranked queries by frequency throughout the day, while others vary dramatically according to query volume. Referring back to Figure 6 and Figure 9, uniqueness of queries in particular categories does not appear to be correlated with the number of queries in their respective category lists, the proportion of the query stream they represent, or the number of distinct queries they match.

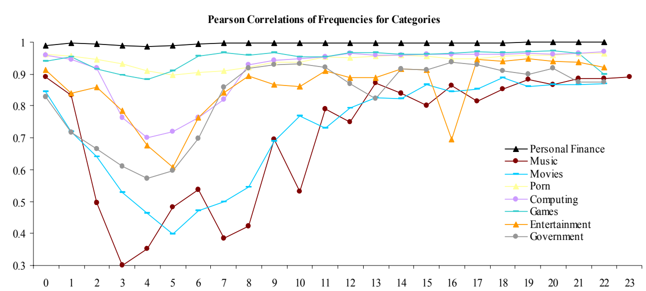

Figure 12

This type of data is potentially of great use to query caching algorithms. For example, if it is known a priori that queries for certain categories are similarly ranked throughout the day, they can be given higher priority in a query-caching scheme. Similarly, queries in categories whose rankings change vastly over time might be given low caching priority.

5. CONCLUSIONS AND FUTURE WORK

This study focuses on investigating the nature of changes in the query stream of a very large search service over time. Understanding how users' queries change over time is critical to developing effective, efficient search systems and to engineering representative test sets and evaluations that drive this development. In this study we find trends over time that are stable despite continuing fluctuation in query volume. Although the average query is repeated only twice during any given hour of the day, the total query traffic varies both in magnitude from one hour to the next, and also in degree of overlap and correlation in popularity of the queries that are received. In addition, we also find that the frequency distribution of an hour's worth of queries remains constant throughout the day. Also, at the most general level, we find that query volume is highest and query sets are most stable during peak hours of the day.

This study further investigates changes in the query stream over time by examining the nature of changes in popularity of particular topical categories. For this we use a set of topical categories created by human editors that represents approximately 13% of the average query traffic. We show that popularity of some of these categories fluctuates considerably while other categories remain relatively stable over the hours in a day. Additionally, we show that the overlap and correlation in popularity of the queries within each topical category varies quite differently over the course of the day.

Extending this analysis to investigate changes in the very rare queries not often matched by our category lists would provide insight into whether those are changing similarly to more popular queries. One method for approaching this might be to incorporate automatic query classification methods to extend our basic lists

This study is the gateway to a large and diverse body of future work. Integrating this knowledge of Circadian changes in the query stream by category will likely yield improved query disambiguation, query caching, and load balancing algorithms.

6. BIBLIOGRAPHY

[1] Beitzel, S., Jensen, E., Chowdhury, A., and Grossman, D. Using Titles and Category Names from Editor-driven Taxonomies for Automatic Evaluation. In Proceedings of *CIKM'03* (New Orleans, LA, November, 2003), ACM Press.

[2] Broder, A. A Taxonomy of Web Search. *SIGIR Forum* 36(2) (Fall, 2002).

[3] Chowdhury, A., G. Pass. "Operational Requirements for Scalable Search Systems", In Proceedings of *CIKM'03* (New Orleans, LA, November 2003), ACM Press.

[4] Eastman, C., B. Jansen, "Coverage, Relevance, and Ranking: The Impact of Query Operators on Web Search Engine Results", *ACM Transactions on Information Systems*, Vol. 21, No. 4, October 2003, Pages 383–411.

[5] Eiron, N., K. McCurley. "Analysis of Anchor Text for Web Search", In Proceedings of *SIGIR'03* (Toronto, Canada, July 2003), ACM Press.

[6] Hawking, D., Craswell, N., and Griffiths, K. Which Search Engine is Best at Finding Online Services? In Proceedings of *WWW10* (Hong Kong, May 2001), Posters. Actual poster available as http://pigfish.vic.cmis.csiro.au/~nickc/pubs/www10actualposter.pdf

[7] Jansen, B. and Pooch, U. A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology* 52(3), 235-246, 2001.

[8] Jansen, B., Spink, A., and Saracevic, T. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information Processing and Management*, 36(2) (2000), 207-227.

[9] Jansen, B.J., Goodrum, A., Spink, A. Searching for multimedia: video, audio, and image Web queries. *World Wide Web* 3(4), 2000.

[10] Lawrence, S. and Giles, C.L. Searching the World Wide Web. *Science* 280(5360), 98-100, 1998.

[11] Lempel, R. and Moran, S. Predictive caching and prefetching of query results in search engines. In Proceedings of *WWW12* (Budapest, May 2003).

[12] Markatos, E.P. On Caching Search Engine Query Results. In the Proceedings of the *5th International Web Caching and Content Delivery Workshop*, May 2000.

[13] Raghavan, V. and Sever, H. On the Reuse of Past Optimal Queries. In Proc. of the *1995 SIGIR Conference*, 344-350, Seattle, WA, July 1995.

[14] Ross, N. and Wolfram, D. End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine. *Journal of the American Society for Information Science* 51(10), 949-958, 2000.

[15] Saraiva, P., Moura, E., Ziviani, N., Meira, W., Fonseca, R., Riberio-Neto, B. Rank-preserving two-level caching for scalable search engines. In Proc. of the *24th SIGIR Conference*, 51-58, New Orleans, LA, September, 2001.

[16] Silverstein, C., Henzinger, M., Marais, H., and Moricz, M. Analysis of a very large web search engine query log. *SIGIR Forum* 33(1) (Fall, 1999), 6-12.

[17] Spink, A., Ozmutlu, S., Ozmutlu, H.C., and Jansen, B.J. U.S. versus European web searching trends. *SIGIR Forum* 36(2), 32-38, 2002.

[18] Spink, A., Jansen, B.J., Wolfram, D., and Saracevic, T. From E-sex to e-commerce: Web search changes. *IEEE Computer*, 35(3), 107-109, 2002.

[19] Spink, A., Wolfram, D., Jansen, B.J. and Saracevic, T. Searching the Web: The Public and Their Queries. *Journal of the American Society of Information Science* 53(2), 226-234, 2001.

[20] Spink, A., Jansen, B.J., and Saracevic, T. Vox populi: The public searching of the web. *Journal of the American Society of Information Science* 52 (12), 1073-1074, 2001.

[21] Spink, A., Jansen, B.J., and Ozmultu, H.C. Use of query reformulation and relevance feedback by Excite users. *Internet Research: Electronic Networking Applications and Policy* 10 (4), 2000.

[22] Sullivan, D. Searches Per Day. *Search Engine Watch*, February, 2003. http://searchenginewatch.com/reports/article.php/2156461

[23] Wang, P., Berry, M., and Yang, Y. Mining longitudinal web queries: Trends and patterns. *Journal of the American Society for Information Science and Technology* 54(8), 743-758, June 2003.

[24] J. Wen, J. Nie, H. Zhang "Query Clustering using User Logs" *ACM Transactions on Information Systems*, Vol. 20, No. 1, January 2002, pp59-81.

[25] Wolfram, D., H. Xie, "Subject categorization of query terms for exploring Web users' search interests", *Journal of the American Society for Information Science*, v.53 n.8, p.617-630, June 2002.

[26] Xie, Y., O'Hallaron, D. Locality in Search Engine Queries and Its Implications for Caching. *Infocom* 2002.