Studies of the Onset and Persistence of Medical Concerns in Search Logs

Ryen W. White
Microsoft Research
One Microsoft Way
Redmond, WA 98052
ryenw@microsoft.com

Eric Horvitz
Microsoft Research
One Microsoft Way
Redmond, WA 98052
horvitz@microsoft.com

ABSTRACT

The Web provides a wealth of information about medical symptoms and disorders. Although this content is often valuable to consumers, studies have found that interaction with Web content may heighten anxiety and stimulate healthcare utilization. We present a longitudinal log-based study of medical search and browsing behavior on the Web. We characterize how users focus on particular medical concerns and how concerns persist and influence future behavior, including changes in focus of attention in searching and browsing for health information. We build and evaluate models that predict transitions from searches on symptoms to searches on health conditions, and escalations from symptoms to serious illnesses. We study the influence that the prior onset of concerns may have on future behavior, including sudden shifts back to searching on the concern amidst other searches. Our findings have implications for refining Web search and retrieval to support people pursuing diagnostic information.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process, information filtering*

Keywords

Health search; medical search; cyberchondria.

1. INTRODUCTION

People often turn to the Web to find information about medical concerns. A recent study found that 80% of Web users have performed online medical searches [12]. Framed by the popularity of health-related search, we seek to better understand how users find medical information online, with the goal of developing tools to better support search for medical information. Researchers have investigated the search behavior of medical domain experts (e.g., [5]) and the reliability of online medical content (e.g., [11]). Search log data provides unique insight into the natural search behaviors of search engine users, and a number of studies have used that type of data to investigate how users find medical information. Studies to date have largely focused on medical search behavior within one session and examined issues such as the escalation of basic symptoms to more serious concerns [24] or decisions about when to seek a consultation with a health professional [25]. However, medical searches span multiple sessions and are connected over time, as people monitor symptoms and adapt their lines of inquiry based on exposure to information found on the Web and other sources (e.g., professional diagnoses) [2].

SIGIR'12, August 12–16, 2012, Portland, Oregon, USA.

We explore how medical concerns emerge over time and how these concerns persist after their onset. Within-session analysis is limited in this context as it may not capture the behavior preceding the onset of searches for a condition, meaning that we may not observe changes from *symptoms* (i.e., an observed departure from normal function or feeling such as headache or chest pain) to *conditions* (i.e., a disease or disorder such as brain tumor or heart attack) within session boundaries. Indeed as we show in our analysis of large-scale search logs, about 60% of medical search sessions start directly with a condition query[1]. The relationship between log-based signs of the onset of concerns and longer-term search behavior following a search diagnostic on the implications of a symptom may or may not be tightly linked; search on conditions may be initiated by external factors such as diagnosis provided by a medical professional or prior searching (in the log data in our study, around 95% of sessions originating with a condition had medical searches in previous sessions). Regardless, a better understanding of the behavior may provide valuable clues about searchers' medical trajectories that search engines might use to help searchers (e.g., by predicting emerging concerns and personalizing the search experience).

We analyze search and browsing data gathered from a sample of approximately 170 thousand consenting toolbar users over a three-month period. Contributions extending from the analysis include:

- Characterizing how health-related concerns emerge in search activity over time, including key aspects of this process such as escalations in the severity of conditions searched.
- Analyzing aspects of the persistence of concerns after the onset, including associated interruptions to future search tasks and changes in the searcher's focus of attention thereafter.
- Construction and evaluation of machine-learned models to predict medical condition onset, escalations, and likely interruptions in future sessions given a user's search history.
- Proposal of design implications based on our findings that can help inform the development of search tools to monitor and proactively support online medical searching.

Through improved understanding and modeling of emerging concerns, we can help searchers make better decisions on their own health seeking and guide engagement with medical professionals.

The remainder of this paper is structured as follows. Section 2 describes related work in such areas as medical search and long-term user profiling. Section 3 describes the data and how we use it to build the long-term user profiles necessary for our study. In Section 4, we present findings of our analysis characterizing the onset and persistence of medical concerns over time. Section 5 focuses on predicting: (i) the onset on concerns, (ii) escalations to searches about serious illnesses, and (iii) interruptions of future

[1] For 81% of search sessions containing terms indicating a specific medical condition, the session started with the same condition.

searches on other topics with returns to the healthcare concern at focus. Section 6 discusses our findings, their implications, and opportunities for future study. We conclude in Section 7.

2. RELATED WORK

Web interaction logs have been used previously to study medical Web search behavior. Bhavnani *et al*. [6] demonstrated that Web page term co-occurrence of medical symptoms and disorders can reasonably predict the degree of influence on search behavior. Spink *et al*. [17] characterized healthcare-related queries issued to Web search engines and showed that users were gradually shifting from general-purpose search engines to specialized Web sites for medical- and health-related queries. White and Horvitz [24] used a log-based methodology to study escalations in medical concerns during Web search (shifts from search on common symptoms to searches on serious ailments) a behavior that they termed *cyberchondria*. Their work highlighted the potential influence of several biases of judgment demonstrated by people and search engines themselves, including base-rate neglect and availability. In a follow-up study, White and Horvitz [25] examined the occurrence of escalation from symptoms to searches for local medical professionals and facilities such as nearby hospitals. Cartright *et al*. [8] studied differences in search behaviors associated with diagnosis vs. more general health-related information seeking. They decomposed health information seeking into evidence-based (search on relevance of signs and symptoms) and hypothesis-based (search on conditions and treatments), and studied how medical foci evolve during exploratory health search sessions.

The research described above largely focused on medical search behavior within a single session. Several pieces of prior work explore the pursuit of health information across multiple sessions. White and Horvitz [24] framed directions with longer-term analyses including a study of the recurrence of concerns over multiple sessions and interruptions of non-medical searches with return to the pursuit of a health concern. They showed that 13.5% of the medical searchers in the logs they studied exhibited recurrence in medical searches, mainly for symptoms (50.8% of users reissued queries for the same symptoms over multiple sessions). In other cross-session research, White and Horvitz [25] show how using evidence drawn from multiple search sessions can enhance the performance of a classifier that predicts the search for local healthcare resources. Another study of medical search behavior over time was conducted by Ayers and Kronenfeld [2] who used Web log data in a multiple regression analysis exploring the relationship between chronic medical conditions and frequency of Web use, as well as changes in health behavior associated with Web usage. Their findings suggest that Web use was influenced not only by the presence of one particular chronic illness but by the total number of chronic conditions. They also found that the more often someone uses the Web as a source of health information, the more likely they will alter their health behavior based on the information they encounter online.

Prior research has used long-term histories more generally to improve search and recommendation effectiveness, typically learning a user's profile from either browsing or search history. White *et al*. [21] investigated the utility of different contextual sources and their combinations for a URL recommendation task. They found that models constructed from historic data for each user most accurately predict users' longer-term future interests. Piwowarski and Zaragoza [15] explored three different predictive click models using historical and social contexts, in a Web search setting trying to predict relationships between queries and documents. Teevan *et al.* [20] constructed user profiles from indexed desktop documents and showed that this information could be used to re-rank search results and improve relevance for individuals. Matthijs and Radlinski [14] constructed user profiles using users' browsing histories, and evaluated their approach using an interleaving methodology. Tan *et al*. [18] focused only on the most relevant prior queries and constructed language models for this task. Although this work has been carried out in the context of general Web search, it has special applicability to medical search.

The medical community has studied the effects of health anxiety disorders, including hypochondriasis, over time [1], but not in the context of Web search. Health anxiety is problematic since it is often maladaptive (i.e., out of proportion with the degree of medical risk) and amplified by the fact that those affected are often undiscerning about the source of their medical information [19]. Such anxiety usually persists even after an evaluation by a physician and reassurance that concerns about symptoms lack an underlying medical basis. Beyond interactions with medical professionals, patients' health concerns may manifest in other ways such as search behavior, an assertion supported by [23], which showed that those whom self-identified as hypochondriacs searched more often for health information than average Web searchers.

Despite the popularity of medical Web content, there is a need for improved information services to the medical information consumer. Cline and Haynes [7] present a review of work in this area that suggests that public health professionals should be concerned about online health seeking, consider potential benefits, synthesize quality concerns, and identify criteria for evaluating online health information. Eysenbach and Kohler [11] used focus groups and naturalistic observation to study users attempting assigned search tasks on the Web. They found that the credibility of Web sites (in terms of source, design, appearance, language used, and ease of use) was important in the focus group setting, but that, in practice, many participants largely ignored source. Some organizations have already started efforts to improve the general quality of medical information on the Web, such as the Health on the Net Foundation [12]. Others have studied enhancing Web pages or search results with tools to support credibility assessment [16]. Eastin and Guinsler [10] showed that an individual's level of health anxiety moderates the relationship between health seeking and healthcare utilization decisions. White and Horvitz [26] demonstrated a relationship between the nature and structure of Web page content and the likelihood that users would escalate their concerns as observed in downstream behavior (e.g., the order in which serious and non-serious explanations for symptoms appeared on the page and the distance between these mentions).

We believe that much more can be learned from studying health information-seeking behaviors over time, and we perform a longitudinal study with that focus. Our work differs from prior research in a number of ways. First, we study the emergence, onset, and the persistence of medical concerns over time as observable in search queries. Second, we leverage a log-based methodology that provides data in the wild from a large sample of consenting users over many search sessions, rather than relying on running user studies, or within-query and within-session log analysis. Finally, we develop predictive models of whether onsets will occur, the nature of the onset, and its persistence, rather than immediate health-related actions such as within-session escalations.

3. BUILDING USER PROFILES

To study long-term medical search behavior, we needed to build long-term search profiles for each user in our log data. We begin by describing the data and profile construction method.

3.1 Data

We use anonymized logs of URLs visited by users who consented to provide data through a widely distributed browser toolbar. We gathered data over a three-month period (January 2011 – March 2011). Log entries included a unique user identifier, a timestamp for each page view, and the URL of the page visited. We excluded intranet and secure (https) URL visits at the source. To remove variability caused by cultural and linguistic variation in search behavior, we only include log entries generated by users in the English-speaking United States locale. From these logs, we extracted search sessions on Google, Yahoo!, and Bing via a similar methodology to White and Drucker [22]. Search sessions comprised queries, clicks on search results, and pages visited during navigation once users left the search engine. Search sessions ended after a period of user inactivity exceeding 30 minutes.

We assigned topical labels to extracted URLs to identify which were medically related. We assigned URLs in our dataset to categories in the Open Directory Project (ODP, dmoz.org) in an automated manner using a content-based classifier, described and evaluated in [4]. The classifier employs logistic regression to predict the ODP category for a Web page, resulting in a label for each page such as $Health/Medicine$. To lessen the impact of small differences in assigned labels, we use only 219 categories at the top two levels of the ODP hierarchy. The findings in [4] revealed that, when optimized for the score in each category, the content-based classifier has a micro-average F1 of 0.60, which we believed was sufficient for our purposes.

3.2 Long-Term Profile Generation

To identify searchers showing evidence of health-seeking intent, we constructed profiles for a randomly selected subset of users who had visited at least one URL labeled with the $Health$ category of the ODP[2]. We removed outlier users (approximately the most active 1% of users) who visited large numbers of URLs in the study duration (as likely traffic generated by robotic systems). The resultant set of approximately 170,000 users, conducted over 25 million search sessions and visited over one billion URLs over the three-month duration of study. The average number of browsing sessions per user was around 240 (of which around 150 were search sessions), spanning on average 76 days per participant.

In addition to labeling URL content using a text classifier, we also automatically labeled the search queries issued by users based on whether they contain symptoms (e.g., headache), benign explanations (e.g., caffeine withdrawal), serious illnesses (e.g., brain tumor), or some combination of these types of medical concern. To perform the labeling, we leveraged lists of relatively common symptoms, as well as benign and more serious illnesses. Specifically, we used the following sources for each class of concern:

- *Symptoms*: Symptom list from the Merck medical dictionary. The same list of symptoms has been used in previous analysis of search behavior [8] to identify health-related sessions.
- *Benign explanations*: List of commonly occurring conditions, as defined in [24] and used in that study for a log-based analysis of online search behavior. The wordlist is based on the *International Classification of Diseases 10th Edition* (ICD-10) published by the World Health Organization.
- *Serious illnesses*: List of serious conditions defined in [24].

In the log-centric analysis, we also used synonyms of symptoms and conditions to increase coverage (e.g., including "tiredness" in addition to "fatigue"). Synonyms for each symptom or condition were identified via a two-step walk on the search engine click graph using an approach similar to that described in [3]. The automatically-generated lists of synonyms were reviewed by the authors to remove erroneous list entries (e.g., all astrology-related synonyms were removed for the serious illness "cancer"). To count as a match on one or more of the concern types, a query needed to be an exact match against the concern or one of its synonyms as generated using the walk described above. We avoided substring matches to ensure high precision in labeling queries.

The user profiles conducted as described in this section were used to pursue answers to three research questions:

Q1: What are some key characteristics of the long-term profiles generated to study health seeking over time?

Q2: What are characteristics of how health-related concerns emerge in search activity *over time*, including key aspects of this process such as escalations in the severity of conditions searched?

Q3: How do searches for the condition persist following evidence of an initial onset of concern, in terms of related medical searches and specifics such as interruptions in future search sessions and changes in searchers' focus of attention?

Answers to these questions could provide valuable insights to search system designers.

Table 1. Summary statistics of long-term medical profiles. Mean and standard deviation (SD) are reported.

Statistic (n=169,513)	Mean	SD
# queries	1076.31	1112.32
# medical queries (click)	36.67	60.29
% queries medical (click)	3.4%	4.3%
# medical queries (wordlist)	10.31	17.90
# queries with serious illnesses	4.60	8.68
# unique serious illnesses	1.69	1.60
# queries with benign explanations	1.86	5.81
# unique benign explanations	0.70	1.05
# queries with symptoms	4.15	11.39
# unique symptoms	1.23	1.75
# URLs	5738.76	5968.84
# medical URLs	190.65	317.53
% URLs that are medical	3.9%	4.6%
# search sessions	153.40	91.55
# medical sessions	37.04	32.05

4. CHARACTERIZING BEHAVIORS

We divide this section into three parts: (i) we provide an overview of medical search behavior in the data set (Section 4.1), (ii) we characterize aspects of the onset of medical conditions (Section 4.2), and (iii) we analyze the persistence of the condition searching following the condition onset (Section 4.3).

4.1 Medical Searching Over Time (Q1)

We begin with an overview of medical search behavior over time, averaged on a per-user basis. Table 1 lists summary statistics over the available log data in the three-month period studied. Note that for the purposes of this analysis, a *medical session* is defined as a search session where the user has issued a medical search query or visited a URL labeled with the $Health$ category of the ODP (or a sub-category). A *medical query* is defined in two ways: *wordlist* –

[2]We excluded URLs in the $Health/Animals$ category.

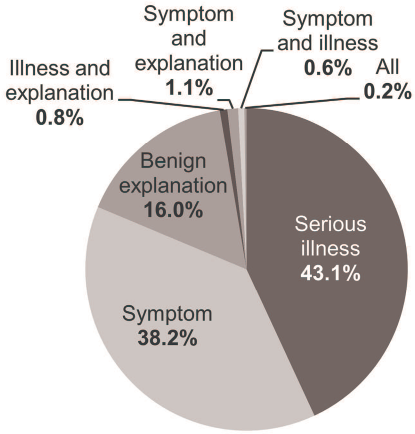

Figure 1. Distribution of medical queries per concern type. Queries are labeled using the wordlist method.

query contains at least one of the symptoms, benign explanations, or serious illnesses defined earlier, or *click* – a query resulting in a search engine result page (SERP) click on a page labeled in the *Health* category of the ODP. The click method has better coverage since it does not depend on string matching. However, since much of our analysis uses the wordlist method (since it is more precise and already broken into conditions, symptoms, etc.), we also report statistics on that.

From the findings presented in Table 1, we can see the average user in our dataset issued 10−15 queries per day. Of those queries, around 3% are medically related. It is worth noting the high standard deviation values for some of the medical statistics, such as the *# medical queries*. As noted previously [23], there are users who may be more likely to search for medical information (e.g., those affected by health anxiety). We retained these and other outliers to more fully represent the distribution of Web searchers.

Table 1 also shows that users are querying more often for symptoms and serious illnesses than common, benign explanations. It is important to understand the composition of medical queries that users are issuing since it has direct implications for the type of support that search engines should offer. With that in mind, we performed further analysis of the queries in medical sessions.

4.1.1 Query Composition

Figure 1 shows the fraction of medical queries (from those that could be labeled according to our wordlist method) that contain each of the seven possible combinations of the three concern types. The figure shows that searching for a single concern type is most common, with 97.3% of labeled medical queries containing only a single *type* of concern (and 94.6% of those queries containing only a single concern). Serious illnesses and symptoms are the most popular concern types that searchers pursue. Interestingly, and somewhat concerningly, searches for serious illnesses occur around two-and-a-half times as often as searches for benign explanations, even though the benign explanations of symptoms are typically much more likely given prevalence statistics.

4.1.2 Symptoms and Conditions

Turning our attention to the role of symptoms and conditions in the medical sessions, we found that 80.7% of sessions that contain a medical condition (serious or benign), start with that condition immediately. In contrast only, 49.8% of the sessions containing a medical symptom start with a symptom. If we compute the percentages across all medical sessions, regardless of whether they contain a medical symptom, the differences between symptoms and conditions are amplified. Overall, 17.6% of medical sessions start with a symptom and 57.8% of medical sessions start with a condition. As suggested in previous work [8], searchers may begin with a hypothesis and refine that based on evidence gathered during the session. The longitudinal methodology provides us with insight into user behavior before the session that can help us to better understand how these concerns first emerge.

4.2 Condition Onset (Q2)

We now explore the search behavior of users prior to the first occurrence of a particular condition of interest across the three-month study duration. We refer to the first query that the user issues related to this condition as the *condition onset*. In identifying the onset, we randomly selected one condition per user from that user's long-term profile. Although the onset is the first instance of the selected condition in the user's long-term search history, their pre-onset history can contain *other* conditions that may have been searched for before the onset. Allowing users to explore multiple conditions more accurately reflects real Web search behavior than, say, limiting our analysis to the first instance of *any* condition in the long-term profiles. That said, this may not have been hugely impactful since 83% searched for at most one other condition prior to onset (61% searched for none).

For each randomly selected onset condition, we extracted all queries and URL visits before its onset, extending back to the beginning of the available search logs for that user. We also extracted the behavior for that user after the condition onset, and we use that to study aspects of the estimated impact of the condition on future search behavior, as described later in the paper. The extraction process yielded 158,917 pre-onset histories, one per user. We begin by studying the search behavior in these histories.

4.2.1 Pre-Onset Search Behaviors

We look at a number of variables capturing users' medical search behavior. From pre-onset behavior, we computed the following:

- *Number of queries*
- *Number of URLs*: Number of non-SERP URLs visited.
- *Number of medical queries*: Number of medical queries, defined using the wordlist and click methods from earlier.
- *Number of medical URLs*: Number of non-SERP URLs that are labeled as being in the *Health* category of the ODP.
- *Fraction of URLs medical*: Fraction of all non-SERP URLs that are labeled as being medical.
- *Total time to onset*: Total timespan to the onset of the condition from the start of the log data for that user.
- *Total time online*: Total time that user spends searching online (sum of all dwell time for pages in search sessions).
- *Fraction time online*: Fraction of the time between the start of the log data and the onset that the user spends online.
- *Fraction online time on medical URLs*
- *Number of search sessions*
- *Number of medical search sessions*

We also compute the same four features for the symptoms and conditions observed in the pre-onset history, e.g., for symptoms:

- *Number of queries with symptoms*
- *Number of unique symptoms*
- *Symptom persistence*: Time for which a symptom persists, computed as timespan between its first and last occurrence.
- *Number of symptom sessions*: Number of sessions that contained at least one query with a symptom.

The mean (M) and standard deviation (SD) pre-onset variable values for all conditions are reported in the *All* column in Table 3. There were about 50 days of user behavior in our logs on average before the onset of the condition of interest. Users performed around 20 medical search sessions in that time, spent 8% of their total time searching online and they spent around 3% of their online time examining medical content. Users generally focused on a single symptom and a single condition in the pre-onset phase. Symptom searches persisted for around four days on average, whereas condition searches persisted for around six days.

To understand the relatedness between the symptoms observed in user search histories and the onset conditions, we used the serious-illness→symptom and common-explanation→symptom mappings from a prior study by White and Horvitz [24]. In that study, the authors mapped 12 commonly occurring symptoms (e.g., chest pain, dizziness) to one or more of over 100 conditions based on a review of medical literature. In this study, we first extracted all onset-condition→symptom pairs from each user's search history, computed synonyms for each using the method described earlier, and filtered to retain pairs with one of the 12 common symptoms from [24]. In total, 79.5% of the filtered condition-symptom pairs observed for a user comprised a correct mapping from a symptom to a condition likely to cause it. The high match shows that across many sessions, the symptoms and the onset-condition are strongly related; even if it takes time for queries mentioning the condition directly to be visible in the query stream, there are indications that its emergence extends back over time.

4.2.2 *Effects of Other Factors*

In addition to behavior over all onsets, we also studied the effect of two marginalizations: condition severity and transition type.

4.2.2.1 *Condition Severity*

From all of pre-onsets available, we selected those where the onset condition was a serious illness or a benign explanation; the 7,104 onsets based on queries containing both severity types were removed for this marginalization analysis. The feature values for each of the classes are reported in third and fourth data columns of Table 3. Given the large sample sizes, all differences between severity types across all variables were significant using one-way multivariate analyses of variance (MANOVA) ($F(2,310727) = 5.81$, $p < .01$). All paired differences for each variable between the severity types were also significant at $p < .013$ using Tukey post-hoc testing. There are some noteworthy differences seen with searchers starting with benign versus serious conditions. We found that the *% medical queries*, *# symptom queries*, *# condition queries*, and *# medical queries* features show that the onset of benign explanations was typically preceded by more medical searching than preceded serious illnesses and All (e.g., *% medical queries* is 4% for benign explanations but only 3.2% and 3% for All and serious illnesses respectively). One explanation is that Web search is contributing as a causal influence (rather than just reflecting user concerns), and that these users are better informed from reading more Web pages, and hence able to make better decisions about the condition(s) to which they should transition.

4.2.2.2 *Transition Type*

As well as studying the severity of the onset condition, we also examined the effect of the type of transition that the user makes between symptoms in the pre-onset phase and the onset condition. Prior work examined *within-session* escalations in medical concerns [24], where users were observed to transition from searches on common symptoms such as twitching, to serious illnesses such as amyotrophic lateral sclerosis (also referred to as ALS, motor neuron disease, and Lou Gehrig's disease), and within-session non-escalations, where users transition from querying about symptoms to benign explanations, such as muscle strain and anxiety. However, given the large volume of medical sessions that begin with a condition directly (around 60% as reported earlier in the paper), it is important to also consider another transition type: *between-session* escalations and non-escalations, where the onset condition is still related to a preceding symptom, but that symptom occurs outside of the current session. Figure 2 illustrates the two types of escalation. Queries are represented as circles, Web pages by rectangles, and different types of escalation by single-line arrows (e.g., [lump]→[cancer] within a single session).

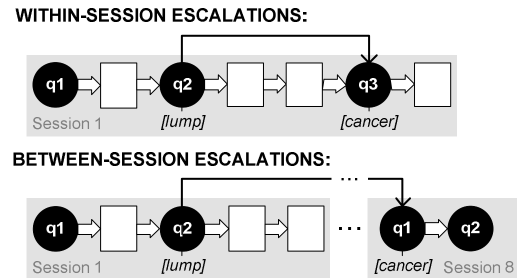

Figure 2. Examples of different escalation types.

To reduce the likelihood that the symptom is unrelated to the condition, we used the serious illness→symptom and benign explanation→symptom mappings described earlier (that restricts the condition-symptom pairs considered to those with a known association) and required that the gap between the symptom and the illness/explanation be at most seven days. We found that within-session escalations occur within three minutes of the symptom query (M=2.57 minutes, SD=4.33 minutes) similar to previous work [24], whereas between-session escalations typically happen within 2–3 days of the symptom (M=2.28 days, SD=1.94 days). We show later that observing escalations leads to better predictive models of the onset condition and post-onset interruptions.

We also computed the values of the features described in the previous section for cases where the onset is a between-session escalation and a between-session non-escalation. The results of this analysis are shown in Table 3. All differences were again significant using a one-way MANOVA ($F(6,321226) = 4.38$, $p < .01$). We can see from Table 3 that escalating and non-escalating users were typically more active searchers on average (*# queries*, *# URLs*), as well as more active in searching for medical information generally (e.g., *% medical queries, # medical sessions*). These users also appear more concerned about symptoms and conditions, as indicated by features such as *# symptom queries* and *# condition sessions*. Focusing on the four transition types, we see that histories ending with a between-session escalation or non-escalation had search behavior that differed most significantly from the overall average. The *Between* columns are highlighted in Table 3 for reference. Between-session escalations and non-escalations may result from long-term concerns which affect medical searching behavior over time, or external diagnosis, where pre-diagnoses concerns are evident in the search log data. Within-session transitions may be a more local phenomenon that is to a lesser extent are reflected in long-term online behaviors.

Table 3 presents statistics over all the pre-onset history, irrespective of where the activity occurred. Since prior work has shown that health behavior changes over time [2], it may also be valuable to understand the temporal dynamics in these features over the course of the pre-onset history and proximal to the onset itself.

4.2.3 *Temporal Dynamics*

We now focus on two temporal aspects of medical searching during the pre-onset behavior: (i) trends over time, and (ii) clusters containing evidence of periodic medical searching.

Table 3. Features of pre-onset search behaviors across all onsets and marginalizations. All differences between groups—All vs. marginalizations, and between and within marginalizations—are statistically significant at $p < .05$. For cells with two entries, first is mean (M) and the second is standard deviation (SD). Bold text denotes features explicitly discussed in the body of the paper.

	Feature	All (n=158,917)		Marginalizations					
				Condition Severity		Transition Type			
		M SD	Gradient of best-fit line	Serious illness (n=117389)	Benign explanation (n=34424)	Escalation		Non-escalation	
						Within (n=1022)	Between (n=8434)	Within (n=439)	Between (n=608)
General health seeking	# queries	611.95 756.86	−0.0140	614.13 740.20	608.79 803.52	**660.93** **672.35**	**778.22** **888.44**	**733.73** **797.96**	**847.14** **880.68**
	# URLs	4148.02 4750.56	−**0.0085**	3985.72 4698.06	3875.80 4887.60	**4298.50** **4897.93**	**4659.43** **5089.18**	**4666.40** **5129.73**	**4863.70** **5181.13**
	# medical queries (wordlist)	5.90 13.11	+**0.1150**	**5.30** **12.12**	**7.69** **15.48**	12.55 32.96	11.64 23.55	13.98 27.07	15.16 33.33
	# medical queries (click)	19.52 41.04	+**0.1230**	**18.02** **39.58**	**24.58** **44.65**	29.10 73.25	35.13 87.25	31.83 63.23	39.06 69.21
	% queries medical (click)	3.2% 3.8%		**3.0%** **4.0%**	**4.0%** **4.2%**	**4.4%** **5.2%**	**4.5%** **4.3%**	**4.3%** **4.8%**	**4.6%** **4.9%**
	Time to onset (days)	**49.50** **26.21**		51.46 25.96	42.84 26.08	50.35 26.24	45.46 26.07	49.57 25.69	46.64 25.11
	Time online (days)	3.89 4.02	−**0.0009**	3.93 4.01	3.70 4.03	4.03 3.87	4.28 4.23	4.46 4.27	4.61 4.27
	% time online (of time to onset)	**7.9%** **8.2%**		7.6% 8.3%	8.6% 9.3%	8.0% 9.2%	9.4% 9.8%	9.0% 9.4%	9.9% 10.3%
	% online time on medical URLs	**2.3%** **2.1%**		2.3% 2.5%	2.4% 2.4%	2.5% 2.9%	2.9% 3.2%	2.6% 3.0%	3.3% 3.8%
	# medical sessions	**21.17** **23.34**	+0.0761	21.09 23.00	21.33 24.50	22.72 27.59	29.23 27.87	24.52 27.27	30.15 26.46
Symptoms	# unique symptoms	**0.63** **1.21**		0.59 1.15	0.72 1.33	**2.08** **1.57**	**2.40** **1.88**	**2.18** **1.94**	**2.64** **2.15**
	# symptom queries	1.98 6.96	+**0.1320**	**1.81** **6.43**	**2.34** **7.58**	**7.38** **11.83**	**9.53** **22.32**	**9.22** **20.34**	**12.38** **22.97**
	Symptom persistence (days)	**3.41** **8.78**		3.33 8.75	3.43 8.55	4.12 10.70	4.76 9.84	4.05 8.90	6.02 9.41
	# symptom sessions	0.74 1.75	+0.0932	0.68 1.61	0.86 1.94	**2.38** **2.41**	**3.09** **3.98**	**2.73** **4.11**	**3.65** **4.76**
Conditions	# unique conditions	**0.74** **1.41**		0.69 1.36	0.89 1.55	**1.09** **1.74**	**1.49** **2.01**	**1.29** **2.01**	**1.64** **2.00**
	# condition queries	1.88 5.68	+**0.1570**	**1.74** **5.36**	**2.27** **6.36**	**3.05** **7.78**	**4.19** **8.09**	**4.26** **10.77**	**5.36** **16.43**
	Condition persistence (days)	**6.04** **12.37**	+0.0400	5.96 12.46	6.13 12.08	5.96 12.39	6.92 13.53	8.06 13.95	5.87 9.55
	# condition sessions	0.94 2.15	+0.0910	0.87 2.05	1.13 2.43	**1.34** **2.50**	**2.01** **3.19**	**1.83** **3.47**	**2.29** **3.85**

4.2.3.1 Trends

For each feature, we computed its daily value (sum or average) across all days in the pre-onset history. If there was no data for a day, then a value of zero was recorded. The result was a time series for each feature value. We then computed the line of best fit for the data and recorded the gradient of the line. If the gradient was positive, then the feature value increased over the time prior to the onset (negative: decreased and zero: constant). The average gradient values for features described in the previous section are shown in Table 3 under the *All* column for applicable variables. In the table, we see that non-medical features such as *# URLs* or *time spent online* are more or less static over the course of the pre-onset phase. In contrast, the features of the medical search behavior, particularly those focused on symptoms and conditions (*# symptom queries*, *# condition queries*) increase over the duration of the pre-onset history, perhaps as the onset condition and related symptoms received a greater focus of attention from the searcher.

As well as computing the average gradient across all conditions, we also studied the gradients by condition to obtain more detailed insight into temporal effects. Figure 3 shows the average gradient values for the *# symptom queries* (which had one of the highest average gradients in Table 3) across all conditions observed in our data occurring 100 or more times. Serious illnesses are marked in dark gray and benign explanations appear in light gray. We can see that benign explanations generally have a lower gradient than the serious illnesses, signifying that there may be more gradual increases in symptom-search frequency related to conditions of that type. Two more noteworthy observations can be made

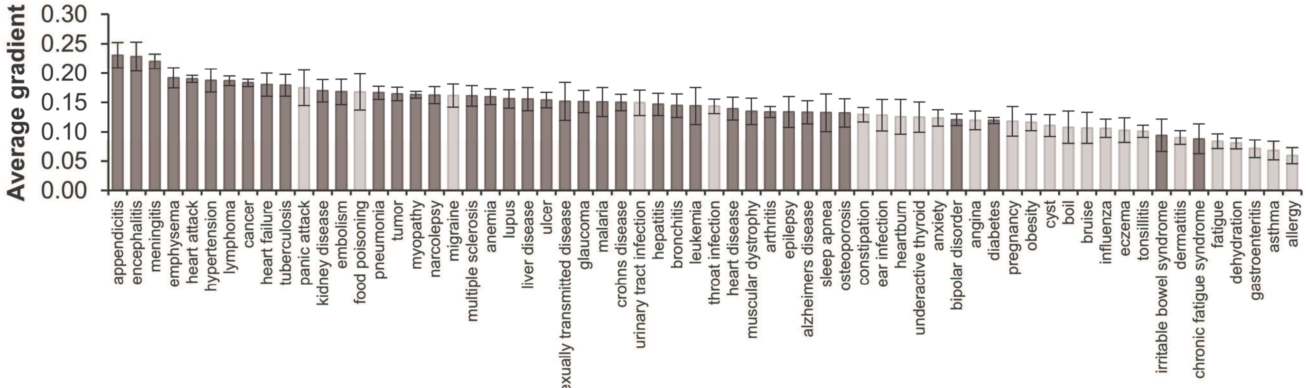

Figure 3. Distribution of gradients of best fit lines for # *symptom queries* (per onset condition). Error bars denote standard error.

from the figure: (i) there is an increase in the acuity of the conditions as average gradient increases and (ii) benign explanations appearing on the left-hand side of the figure are acute conditions such as panic attacks and food poisoning, whereas the serious illnesses on the right-hand side are chronic syndromes and diseases which are stable over time, e.g., diabetes. This analysis shows that we can estimate some important properties of the onset condition by looking at trends in preceding symptom searching. As we will show later, these trends have strong predictive value.

4.2.3.2 Periodicities in Medical Searching

Considering only temporal trajectories tells us little about how the features are distributed temporally. Previous work [24] briefly described the presence of periodicities in medical searching, with periods of no medical search interspersed with clusters of intense activity. Here, we identified such clusters for pre-onset histories.

We define a *cluster* as a sequence of consecutive medical queries with at least one symptom or condition query per day. A gap of a day or more leads to the creation of a new cluster. The average number of clusters per pre-onset history is 2.83 (SD=2.43), the average cluster length is around 2.54 days (SD=2.32 days), and the average periodicity of the clusters is around two weeks (M=14.50 days, SD=12.70 days). To determine whether periodicity changed as the onset approached, we computed the Pearson correlation coefficient (r) between the time between the clusters and the time from the onset. There was a fairly strong positive correlation between these variables (r=0.55) showing that as we move further back in time away from the onset the distance between the clusters increases. In addition, as the time until the condition onset drops, average cluster duration increases (r=−0.64). Trends in both of these measures suggest that the intensity of the medical searching increases as the condition onset nears.

4.2.4 Summary

So far, we have explored aspects of the pre-onset search behavior. Among other things, we showed that: (i) the nature of the condition and the symptom-onset transition may affect medical search behavior, and (ii) that symptom and condition searches increase prior to condition onset (varying by condition type), with an observed increase in the intensity of medical searching. These insights may be useful in predicting outcomes such as condition onsets, something that we explore in more detail later in the paper.

4.3 Persistence of Conditions (Q3)

We now turn attention to post-onset search activity, including important issues such as the amount of time that searchers spend searching related to the onset condition, interruptions to subsequent search activity that are related to that condition, and changes in the conditions that searchers' attend to over time.

4.3.1 Post-Onset Search Behavior

We computed the same features for post-onset behavior as we did for onset (Table 4). We also show the relative percentage change between pre-onset and post-onset behaviors in the last column and features for the onset condition (highlighted), including the number of times the condition recurred post-onset and the total time spent on the onset condition, the sum of the dwell times of all pages on click trails following each instance of a condition query.

The results of this analysis suggest that users are more active medical searchers once they focus on a particular condition. All differences between pre- and post-onset values of the variables in Table 4 are significant (independent t-tests, Bonferroni correction, all $p \leq .002$). We observe a drop in symptom searching, perhaps because they are now focused on a specific condition. The number of unique conditions actually rises because only 39% of users perform condition searching pre-onset; post-onset condition searching is much more common, centered on a single condition.

We now study ways in which the onset might persist in subsequent search behavior: as interruptions to unrelated search tasks and in changes to users' focus of attention over time.

4.3.2 Interruptions

Interruptions refer to a suspension of an ongoing activity and have been studied in detail in the human-computer interaction and psychology communities (e.g., [9]). The analysis of interruptions to search activities associated with persistent health concerns was first introduced in [27]. That research focused on interruptions caused by escalations or health anxiety, and found that interruptions were largely occurred as recurrent searches for symptoms. In this paper, we extend our previous work to consider the all interruptions associated with the onset condition (irrespective of the cause) and the specifically focus on the cost of the interruptions in terms of time users spent and frequency of occurrence in the logs.

To study interruptions in our log-based context, we needed a way to automatically label interruptions so as to be reasonably confident that the user was being interrupted by medical concerns. To do this, we used the following definition: Given a search session containing queries and URLs, an interruption comprises at least one search for the onset condition sandwiched between two series of one or more non-medical searches. Analysis of the post-onset data revealed that 30.3% of users were interrupted in this way. Of those who experienced an interruption, there were 2.39 interrup-

Table 4. Features of post-onset search behaviors. Degree of change from pre-onset behaviors are also shown.

Feature	% or M ±SD	% change fm pre-onset
% URLs medical	4.9%	+88.5
% queries medical	4.2%	+31.3
% online time on medical pages	8.0%	+247.8
# unique symptoms	0.50 ±1.01	−20.6
Symptom persistence (days)	2.46 ±3.42	−27.9
# unique conditions	1.04 ±1.29	+40.5
Condition persistence (days)	7.57 ±12.49	+25.3
# onset condition recurrence	2.72 ±4.79	n/a
Time on onset condition (hours)	1.43 ±2.11	n/a
Onset condition persistence	9.97 ±18.69	n/a
# onset condition sessions	1.78 ±1.67	n/a

tion events and they spent a total of 2 hours 16 minutes searching for the onset condition (SD=2 hours 31 minutes, max=30 hours 48 minutes). Around 6% of search sessions were interrupted and on average 2.68 days (of the on-average 27 post-onset days) contained an interruption per user. It seems that the onset of concern about a condition does have an impact that is in part manifested as interruptions to other future search behavior. Further evidence for which the onset condition may affect post-onset behavior is found by examining what has the user's focus of attention over time.

4.3.3 Focus of Attention

Focus of attention in our context refers to the conditions and/or symptoms that the searcher is targeting during health seeking. Cartright *et al.* [8] studied this concept with a single search session. We expand that research by examining searcher foci post-onset and spanning multiple search sessions, specifically looking at the extent to which searchers focus only on the onset condition.

We explored the post-onset data for evidence of a narrow (single) focus of attention once the onset emerged. In particular, we focus on cases where we observed users searching for at least two unique conditions (N=27,016), including the onset condition, and then after condition onset, *the user no longer performs any symptom searches and only searches for the onset condition*. In analyzing our data, we observe that 27.9% of users exhibited evidence of such narrowing of their attentional focus post-onset. The effect is amplified given particular transition types such as escalations.

Table 5 shows the percentage of users with a narrower focus of attention in their post-onset search behavior, broken down by transition direction and by transition distance. The findings summarized in the table suggest that a singular focus of attention is much more likely following escalations, and in particular, following between-session escalations. This may be caused by the escalation, an accumulation of medical information from the Web, or from real-world sources such as diagnosis or prognosis by a medical professional. If escalations provide clues on the future behavior of users, then they may be useful for predicting search behaviors and in developing tools to proactively support health seeking.

4.3.4 Summary

In this subsection we have studied aspects of the persistence of onset condition and associated behavior. We show that there is a significant increase in health-search behavior following the emergence of the onset condition, as well as other changes such as a drop in symptom searching. There is also evidence of interruptions related to the onset condition, as well as a narrowing of the

Table 5. Percentage of users who exhibit a narrower focus of attention post onset.

Direction	Distance	
	Within-session	Between-session
Escalation	23.8%	32.7%
Non-escalation	12.6%	16.6%

focus of attention following the condition onset. These are examples of the types of insights that can only be made through our novel examination of long-term search behavior.

The paper so far has presented a descriptive analysis of long-term health-seeking behavior on the Web. We believe that such characterizations are valuable to learn how people pursue medical information online, and can benefit search providers and other researchers working in this area. Although insights can be gleaned from such analysis that can assist in the design of search systems, to have direct practical value to search engine users an application is required. We envision that predictive models based on long-term search behavior will play an increasing role in search systems, monitoring search intentions and managing system intervention. We now explore such predictions in more detail.

5. PREDICTING ONSET & PERSISTENCE

We focus on three tasks: (i) whether users will transition from symptom searching to condition searching, (ii) whether users will escalate between-session to a particular condition, and (iii) whether users will be interrupted by the condition post-onset.

5.1 Feature Generation

The features used in the prediction are in four classes: (i) *queries*, (ii) *pages*, (iii) *temporal* − trends, clusters, and persistence of symptoms or conditions over time, and (iv) *onset* − features of the condition onset, used only in the third task to predict interruptions.

Query features: The query length, number of queries, the number and fraction of queries that are medical, the number of symptoms, serious illnesses, and benign explanation queries. Counts of the various escalation and non-escalation types are also included.

Page features: The total number of pages, the number and fraction of pages that are medical, the number of unique medical ODP categories visited, the total time spent in medical pages, the fraction of the total time online that was spent on medical pages.

Temporal features: Total timespan of the pre-onset history, trends in the feature values over time, and cluster features, such as their duration and periodicity. Also included is the persistence of symptoms and conditions over time (time between instances of them).

Onset features (only used for prediction task (iii)): Features of the type of transition, the type of condition, and the distance from the most recent symptom to the onset, if available.

5.2 Methodology and Findings

We constructed classifiers to address each of the three prediction tasks. We built models using logistic regression, shown to be an effective classifier for similar tasks in the online medical search domain [25]. Since the methodology used differed for each task, we describe the method and the findings separately.

5.2.1 Onset

This task involved predicting whether a user will transition from symptom searching to condition searching. In the analysis presented so for in this paper, we know when the onset has occurred. By that point, it might be too late for a search engine to intervene, and being able to predict whether the onset of a condition will occur in the future could be useful. To do this, we required training data containing all features minus anything condition-related (since we wanted to predict condition presence) and a label indicating whether the user transitioned to a condition in the future.

To construct the training set, we started at the beginning of the available log data for each user and determined whether they had transitioned to a condition search at some point in the future. If the user had transitioned, then we took the time of the first condition onset (which was on average 45 days from the beginning of the logs, denoted as M_C), subtracted a random number of days (between one and seven) from that time (so that we were not always making the prediction immediately before the onset), and used data in the resultant timespan to create features. For users who did not transition to a condition at any point in their log data, we used M_C and subtracted a random number of days (from 1−7 days) from that data. This gave us positive examples (search history→onset) and negative examples (search history→no onset), one example per user. We selected 5000 examples with an equal number of positive and negatives, such that the accuracy of a marginal model that always predicted positive was 50%. We measured classifier accuracy and used ten-fold cross validation.

The findings of the experiments showed that we were able to predict whether a user would transition to a condition with an accuracy of 75.5%, averaged across 100 experimental runs. Important features included the volume of medical content viewed. Figure 4 shows the receiver-operator characteristic (ROC) curve for the onset prediction task averaged across all runs. Also shown in the figure are the curves for the other tasks, which we now describe. Note that all models perform significantly better than the marginal model using t-tests on the 100 runs ($t(198) \geq 3.34$, $p < .01$).

5.2.2 Escalations

In the analysis presented earlier in the paper, we showed that between session escalations have the potential to cause lasting long-term impact on search behavior. Therefore, we investigated the prediction of between-session escalation given prior search behavior. The methodology employed is similar to that for the onset prediction task and the classifier only had access to symptom features. The positive examples resulted in between-session escalations and the negative examples did not. Once again, we selected 5000 examples for training and test and equally split positive examples (search history→between-session escalation) and negative examples (search history→no between-session escalation), and we used ten-fold cross validation. We attain 76.4% accuracy, and the most useful features were temporal (especially the symptom gradients) and unique symptom query counts.

5.2.3 Interruptions

The final task involves predicting whether a future interruption is going to occur given that we have already observed the onset of a particular condition. Interruptions are defined as in Section 4.3.2. For this task, we can use all features preceding the onset and also the features of the onset itself, such as the condition type (e.g., serious illness or benign explanation), the transition (e.g., escalation or non-escalation), and the time taken to perform the transition. As before, we selected 5000 random examples, with an equal number of positives (search history+onset→interruption) and negatives (search history+onset→no interruption).

The results of our analysis show that we were able to attain a prediction accuracy of 80.6% for this task. The features that were most valuable in making the prediction were related to the nature of the onset, suggesting that with knowledge of those features (and no long-term history of search behavior) we would be able to

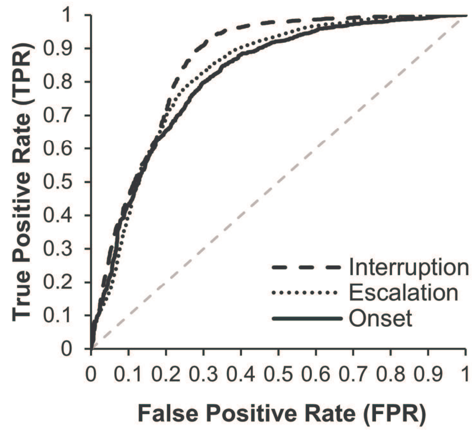

Figure 4. ROC curves for each prediction task. Performance of the marginal model shown as dotted line.

predict whether the user would be interrupted in their future searching. Indeed, we experimented with only restricting the features used to those in the onset class and found that we were able to predict with 71.2% accuracy that the user would be interrupted at some later date by a return to the onset condition.

The models and experiments described in this section demonstrate how key aspects of the condition onset and its estimated impact can be predicted with good accuracy. As we discuss in the next section, accurate predictive modeling capabilities afford a range of possibilities for search engines to better help medical searchers.

6. DISCUSSION AND IMPLICATIONS

We studied the onset and persistence of medical concerns over time as observed from search log data. This is the first study of its kind to focus on long-term medical search. During this study we explored features of the pre-onset search behavior and investigated trends in the feature values and changes in their intensity over the pre-onset phase. Among other things, we found that:

- Many pre-onset symptoms were related to the onset condition, even though they could be issued weeks beforehand.
- Symptom and condition searching increases with proximity to the appearance of evidence of the onset of interest in a specific condition, and the trends are affected by condition acuteness.
- The nature of the onset, measured in this study in terms of severity and transition type (escalations and non-escalations) affect pre-onset search behavior.
- Medical search behavior post-onset differs from pre-onset, e.g., symptom searching drops, condition searches increase.
- Interruptions and a narrowing of the focus of attention succeeding and related to the onset condition occur frequently.

We also showed that we can predict several aspects of the medical search process, such as condition onset, between-session escalation, and related interruptions to future search activities.

Several refinements of search are suggested by the results. On one, we can foresee the use of onset occurrence as a cue for a search engine to leverage that user's previous clicks for related-symptom queries as implicit feedback to enhance the current ranking (similar to [18]). If the search engine *predicts* the onset it could also use the set of symptoms previously searched to estimate the condition that the user is likely to transition to and adjust the search experience based, for example, on severity or incidence rates. The between-session escalation predictor could be used to inform the user *before* they escalate. If the engine can reliably estimate that the user is heading toward an escalation, it could intervene and help the user make an informed decision about what condition to focus on, perhaps offering a set of common benign explanations instead, or more generally, prior and posterior probabilities of serious (rarer) illnesses, conditioning the posterior probabilities on key demographics (e.g., age and gender) and related symptoms, family history, etc. Alternatively, since users may lose time task switching during interruptions [9], the engine could anticipate future interruptions given an onset and proactively gather other information to lessen the lag time when the user returns to a condition search, or provide special support to simplify task switching. The challenge lies in leveraging predictions in a way that respects the highly sensitive nature of medical information, does not overstate their role or their value, and guides users to qualified medical professionals where appropriate.

Although our findings are promising, we acknowledge several limitations of our research. Even though we have a user identifier, it may not actually be a single user who is performing the searches, especially for shared machines. We also did not filter out medical professionals' search data, which is likely to differ from novices [5]. By focusing on a single onset condition for each user we could not dive more deeply into key structural aspects of medical concerns over time, such as transitions between concerns, or how searching and browsing activity associated with common and less common symptoms, benign versus serious disorders, etc. influences these transitions. The association between pre-onset behavior and the onset condition could also have been made clearer by requiring that the onset condition occur within a time window of initial searches for medical concerns; discarding users from our analysis who did not meet this criterion. Also, we did not seek to distinguish the behavior of people with (potentially inappropriate) health anxieties and those wrestling with a diagnosed or yet to be diagnosed serious illness. User attributes, such as levels of health anxiety [23], may influence search behavior. We studied interruptions and focus-of-attention at scale in a naturalistic setting, but used limited definitions of each. More work is needed to refine those definitions for more accurate detection of interruption, attention, and biases. Finally, we need to work with users directly, in addition to their logs, to understand the medically-relevant events in their lives (e.g., professional diagnoses or prognoses), and study how those events affect their search behavior.

7. CONCLUSIONS

We presented a log-based longitudinal study of medical search behavior on the Web. We characterized aspects of search behavior and temporal changes in that behavior before the onset of a condition, and we showed that there are differences in behavior associated with the nature of the onset condition and the transition to that condition from symptom searching. We also studied post-onset behaviors and found evidence of interruption effects and focus narrowing associated with the onset condition. Search engines could monitor long-term search behavior and provide support in advance of the onset or to lessen its impact on search behavior. To this end, we developed predictive models to accurately estimate when concerns are likely to escalate and the impact that their onset can have. Our findings support the development of tailored support for medical search, to help those turning to the Web for medical matters better manage their health seeking.

REFERENCES

[1] Asmundson, G.J.C., Taylor, S., and Cox, B.J. (2001). *Health Anxiety: Clinical and Research Perspectives on Hypochondriasis and Related Conditions*. Wiley.

[2] Ayers, S. and Kronenfeld. J. (2007). Chronic illness and health-seeking information on the Internet. *Health*, 11(3): 327–347.

[3] Beeferman, D. and Berger, A. (2000). Agglomerative clustering of a search engine query log. *KDD*, 407–416.

[4] Bennett, P., Svore, K. and Dumais, S. (2010). Classification enhanced ranking. *WWW*, 111–120.

[5] Bhavnani, S.K. (2002). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. *SIGCHI*, 610–611.

[6] Bhavnani, S.K. *et al*. (2003). Exploring the distribution of online healthcare information. *SIGCHI*, 816–817.

[7] Cline, R.J. and Haynes, K.M. (2001). Consumer health information seeking on the Internet; the state of the art. *Health Education Research*, 16(6): 671–692.

[8] Cartright, M., White, R.W., and Horvitz, E. (2011). Intentions and attention in exploratory health search. *SIGIR*, 65-74

[9] Czerwinski, M., Horvitz, E., and Wilhite, S. (2004). A diary study of task switching and interruptions. *SIGCHI*, 175–182.

[10] Eastin, M.S. and Guinsler, N.M. (2006). Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. *Cyber. and Behav.*, 9(4): 494–498.

[11] Eysenbach, G. and Kohler, C. (2002). How do consumers search for and appraise health information on the world wide web? qualitative studies using focus groups, usability test, and in-depth interviews. *Brit. Med. J.*, 324: 573–577.

[12] Fox, S. (2011). Health topics: 80% of internet users look for health information online. *Pew Internet & Amer. Life Project*

[13] Health on the Net Foundation. www.hon.ch/HONcode/Pro.

[14] Matthijs, N. and Radlinski, F. (2011). Personalizing web search using long-term browsing history. *WSDM*, 25−34.

[15] Piwowarski, B. and Zaragoza, H. (2007). Predictive user click models based on click-through history. *CIKM*, 175-182.

[16] Schwarz, J. and Morris, M.R. (2011). Augmenting web pages and search results to support credibility assessment. *SIGCHI*, 1245–1254.

[17] Spink, A. *et al*. (2004). A study of medical and health queries to Web search engines. *Health Info. and Lib. J.*, 21: 44–51.

[18] Tan, B., Shen, X., and Zhai, C. (2006). Mining long-term search history to improve search accuracy. *KDD*, 718–723.

[19] Taylor, S. and Asmundson, G.J.C. (2004). *Treating Health Anxiety: A Cognitive-Behavioral Approach*. Guilford Press.

[20] Teevan, J., Dumais, S.T., and Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. *SIGIR*, 449−456.

[21] White, R.W., Bailey, P., and Chen, L. (2009). Predicting user interests from contextual information. *SIGIR*, 363−370.

[22] White, R.W. and Drucker. S.M. (2007). Investigating behavioral variability in web search. *WWW*, 21–30.

[23] White, R.W. and Horvitz, E. (2009). Experiences with web search on medical concerns and self-diagnosis. *AMIA*, 696−700.

[24] White, R.W. and Horvitz, E. (2009). Cyberchondria: studies of the escalation of medical concerns in web search. *TOIS*, 27(4): 23.

[25] White, R.W. and Horvitz, E. (2010). Web to world: predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. *AMIA*, 882–886.

[26] White, R.W. and Horvitz, E. (2010). Predicting escalations of medical queries based on web page structure and content. *SIGIR*, 769–770.