Web to World: Predicting Transitions from Self-Diagnosis to the Pursuit of Local Medical Assistance in Web Search

Ryen W. White, PhD and Eric Horvitz, MD, PhD
Microsoft Research, One Microsoft Way, Redmond, WA 98052
{ryenw, horvitz}@microsoft.com

Abstract

Many people turn to the Web for self-diagnosis and healthcare assessment based on limited knowledge of signs, symptoms, and disorders. Studies of search and browsing for healthcare information have shown that reviewing Web content can lead to escalations from concerns about common, typically benign symptoms to searches on rare and frightening disorders. We explore the potential for the Web to induce costly and potentially unnecessary engagements with health professionals. We present findings on the transition of search on common symptoms to the pursuit of in-world healthcare resources such as nearby physicians and healthcare facilities. We build models that predict the transition from searches on initial common symptoms to queries pursuing local medical expertise, using evidence about a user's stream of queries, the content on reviewed pages, and long-term medical search behaviors. Our findings have implications for reducing costly and unnecessary healthcare resource utilization through refinements of ranking algorithms and search interfaces.

Introduction

Consumers frequently use Web search engines to access health-related information[1] and perform self-diagnosis, where queries about symptoms are input and the rank and information of results are interpreted as diagnostic conclusions.[2] However, increased reliance on the Web does not necessarily lead to lessened dependence on health professionals. Indeed, unnecessary visits with physicians may be triggered by online access to poor-quality health information, as people may experience confusion, uncertainty, and anxiety about the information they encounter.[3] Prior research has demonstrated that Web content can fuel concerns and drive people to visit their general practitioner or medical specialist.[4] A survey of over 500 people on healthcare search behavior found that information obtained from medical searches can affect peoples' concerns, their decisions about when to engage a physician for assistance with diagnosis or therapy, and their overall approach to maintaining their health or the health of a family member.[5] Approximately one quarter of survey respondents reported that they were "put over the threshold to engage with a medical professional based on

Web content." Respondents reported that in most cases (72%), their encounter with a physician eased their concerns. Accurate prediction of when users may seek in-world medical advice based on the review of health information on the Web demonstrates the influence of the Web on the anxiety of consumers and on associated costly and potentially unnecessary healthcare resource utilization.

We have found that we can predict with surprising accuracy whether escalations in medical concerns will occur following a page visit.[6] In this paper, we predict when users' searches about common symptoms will transition to queries that show *healthcare resource utilization intent* (HUI). Examples of HUI queries include *[neurologist in seattle, wa]*, *[evergreen hospital]*, and *[urgent care clinic]*. We believe that such signs of intent to pursue information on nearby physicians and health centers are useful proxies for the intention to engage real-world resources.

To build models that predict transitions to HUI queries, we consider features of the structure and content of visited Web pages, features of the interaction behavior during the session, and user's history of medical search. Accurate predictions of the intent to seek medical assistance may one day enable search systems to alert users to the onset of potentially unnecessary healthcare resource utilization, flag specific pages as possible sources of inappropriate escalation, or down-weight such pages in search result rankings.

Related Work

Researchers have provided a rationale for predicting the effects that Web use may have on the frequency of contact with health professionals.[7,8] They suggest that Web usage has a significant impact on aspects of the physician-patient relationship. If people use interactive features of the Web (e.g., email messaging) to contact health professionals, their total time for conversation with health professionals is likely to increase. Also, if people develop better knowledge about health issues by accessing relevant health information on the Web in advance of visits with physicians, their visits are likely to be more productive. Such opportunities can lead to more efficient and satisfactory physician engagements and may lead patients to contact health professionals more often.[9]

Some researchers have shown a negative relationship between Web usage and healthcare utilization[10] (e.g., the Web enables people to make their own healthcare decisions), or no relationship at all.[3] Eastin and Guinsler[10] investigated the relationship between Web health information seeking and healthcare utilization (e.g., visiting a general practitioner). Their findings suggest that an individual's health anxiety moderates the relationship between online health information seeking and external healthcare utilization. Baker et al.[4] examined the prevalence of Web and email use for healthcare and the effect that Web and email has on users' knowledge of healthcare matters and use of the healthcare system. Using self-reports, Baker and colleagues concluded that people rarely use email to communicate with physicians and that the influence of the Web on healthcare utilization is uncertain.

Benigeri and Pluye[11] showed that exposing people who have no medical training to complex medical terminology may put them at risk of harm from self-diagnosis and self-treatment. Studies have also been performed on the challenges associated with the perusal of medical Web content. Cline and Haynes[12] suggest 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 Köhler[13] reviewed several studies and concluded that health-related Web content is often of poor quality.

Identifying Resource Usage

Log Data and Session Extraction

We studied anonymized logs of URLs visited by people who had consented to provide usage data via a widely-distributed browser toolbar. We gathered data over a period of six months from September 2008 until February 2009 inclusive. Log entries include a user identifier, a timestamp for each page view, and the URL of the page visited. We excluded intranet and secure (https) URL visits before log entries were sent to our server. Only entries generated in the English speaking regions of the United States were included. From these logs, we mined many thousands of search sessions using automated tools described in previous work.[2] Sessions comprise a sequence of time-ordered search queries and Web pages, starting with a query and a search result click, and terminating after 30 minutes of user inactivity.

Pursuit of Local Medical Expertise

In our study, we automatically identified search sessions where users transition from queries on common symptoms to queries that show a pursuit of local medical resources. We considered sessions with queries for one of three symptoms: *chest pain*, *muscle twitches*, and *abdominal pain*. We selected these symptoms because they are common health complaints and occur frequently in search logs.

We identified queries on the symptoms and transitions to HUI queries via lookup of terms stored in predefined lists. Through reviewing medical literature and browsing search log data, the authors (one of whom has medical training) constructed lists of terms or phrases capturing how searchers describe each of the symptoms and lists of terms or phrases on seeking in-world medical resources for each symptom.

HUI queries were identified within logs using terms that provide salient evidence of the pursuit of external resources, including words and phrases that indicate the seeking of contact information for local general practitioners and specialists. We identified sessions containing a query on one of the three symptoms. Within each session, we searched for HUI queries appearing later in the session. We identified HUI queries with terms that reflect: (i) an appropriate medical specialty for the symptom (e.g., *neurologist* for muscle twitches); (ii) a name of a medical resource (e.g., *hospital*, *physician*); (iii) a five-digit US zipcode; or (iv) a US city and state name pair (e.g., *Redmond, Washington*). The list of cities used was drawn from United States census data. The state name qualifier was necessary since many locations have ambiguous names (e.g., the city of Early, Iowa).

We used automated log analysis to identify thousands of candidate Web search sessions with and without HUI queries. For quality assurance, we randomly selected search sessions identified by this automated process and inspected the candidate resource queries and preceding queries to verify the accuracy of both the identification of the HUI query and the accuracy of the link to the initial symptom of interest that had been noted. We found 700 search sessions where queries exhibited a transition from an initial search on one of the three concerning symptoms to a HUI query. Over 4% of search sessions with one of the three symptoms contained an HUI query. Another set comprising 700 sessions with an equivalent distribution of the three symptoms but with no HUI queries were also randomly selected for the prediction task.

For sessions with HUI queries, we generated page features from the Web page immediately preceding the HUI query. For sessions without HUI queries, we generated page features from the page immediately preceding session termination. Figure 1 (overleaf) displays sessions with and without healthcare utilization intent, with q as queries, boxes as pages, and P as the page from which we generate page features.

We now present comparisons on the characteristics of sessions with and without HUI queries.
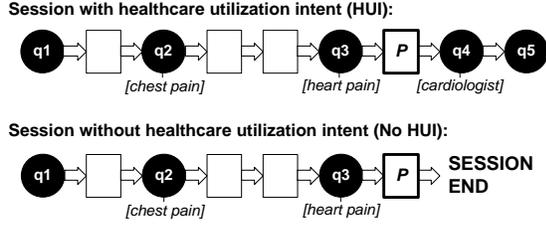
Session with healthcare utilization intent (HUI):

q1 → [ ] → q2 → [ ] → [ ] → q3 → P → q4 → q5

[chest pain]   *[heart pain]*   *[cardiologist]*

Session without healthcare utilization intent (No HUI):

q1 → [ ] → q2 → [ ] → [ ] → q3 → P → **SESSION END**

[chest pain]   *[heart pain]*

Figure 1. Interaction sequences from Web search sessions with and without healthcare utilization intent.

Characteristics of Resource Usage

To characterize sessions that show an intention to seek local healthcare resources, we begin by comparing and contrasting features of search behavior in all search sessions identified as containing a query on one of the three symptoms considered in our study. We held out a set of 100 positive and 100 negative sessions as examples that were not used in the later prediction task, and we probed the predictive value of single features of pages, sessions, and users (e.g., the feature of *SeriousBeforeBenign* is true if page content on a serious illness precedes a benign explanation for the symptom). We compared session features from the first symptom occurrence until the page immediately preceding the first HUI query in HUI sessions or the page preceding session termination in non-HUI sessions. Table 1 shows the average feature values.

Table 1. Features of pages, sessions, and users for sessions with and without healthcare utilization queries.

Features	HUI	No HUI
SeriousBeforeBenign (Page)	59%	48%
IsWebForum (Page)	14%	9%
NumQueries (Session)	4.9	2.9
AvgQueryLength (Session)	4.5	4.1
NumUniqueSymptoms (User)	3.6	2.2
NumResourceQueries (User)	5.5	2.0

The findings illustrate marked differences in features in all classes for sessions with and without HUI. All differences were significant at $p < .05$ using independent measures t-tests. In the next section, we shall explore the construction of predictive models that fuse the evidence from these features and other features that showed less significant differences in prevalence in HUI and non-HUI sessions.

Health resource seeking can take on many forms, ranging from pursuing a consultation with a local specialist to seeking out a local urgent care center. For sessions containing an HUI query, we attempted to establish the class of pursuit. With the use of human labeling of queries identified as having evidence of resource seeking, we observed searchers trying to locate six types of resources: (i) a specific physician (e.g., *[dr chuck na everett]*); (ii) any physician (e.g., *[physicians near 32713]*); (iii) a medical specialist (e.g., *[endocrinologists in chattanooga, tn]*); (iv)

treatment for a symptom or condition (e.g., *[back pain, peoria illinois]*); (v) a specific treatment location (e.g., *[tacoma urgent care]*); and (vi) any treatment location (e.g., *[emergency clinic in sacramento]*). The distribution of these six resource targets across all HUI queries is displayed in Figure 2.
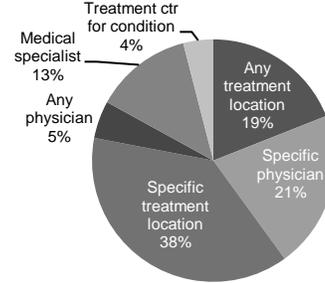

Figure 2. Distribution of resource targets in HUI queries.

The analysis reveals that in seeking external healthcare resources, dominant goals involve locating specific treatment locations or physicians.

Analysis shows that HUI queries begin toward the end of sessions. In 36% of cases, HUI queries are the last query in the session and across all sessions they first occur 75% of the way through the session (where length of session is measured by the number of queries). In cases where additional queries follow the HUI query, manual inspection reveals that search activity falls into one of five categories: (i) refine query in pursuit of particular resource, (ii) request next page of search results, (iii) compare different resources (e.g., compare two medical specialists), (iv) search more about a medical condition, and (v) shift to a different topic. The findings show that searchers primarily refine queries to locate a resource (46% of sessions) or to seek out more information about the symptoms or conditions driving their searches (22%).

In the next section, we use the sample of 700 positive and negative sessions to investigate the feasibility of building a statistical classifier to predict the evolution of queries on basic symptoms to HUI queries.

Predicting Queries with Resource Usage Intention

We constructed classifiers to predict whether a user would next issue an initial HUI query based on a page being viewed. We built models using logistic regression and considered the HUI/non-HUI outcome for each identified session and over 50 features.

Feature Generation

The features used in the prediction fall in three classes: (i) *page*: structure and content of page preceding the query of interest (some of which were useful for predicting medical escalations in earlier work[6]), (ii) *session*: aspects of search interaction in the current search session that fall *between the initial*

concern and the first healthcare utilization request or session termination, and (iii) *user*: aspects of users' medical search interaction history *from the beginning of the logs to the start of current session* (Table 2).

Table 2. Features used in the prediction of HUI queries. Question marks after descriptions signify binary variables.

Page Features
AdsPresent: Advertisements present on page?
NumAdBlocks: Num advertising blocks
SeriousThenBenign: Serious illness for concern appears on the page before a benign explanation for that symptom?
NumWordsToSerious: Num words to first serious illness
NumWordsToBenign: Num words to first benign explanation
NumWordsBetweenSeriousAndBenign: Num words between first serious illness and first benign explanation
SeriousInTitle: Serious illness in page title?
BenignInTitle: Benign explanation in page title?
SeriousInFirstPara: Serious illness in first paragraph?
BenignInFirstPara: Benign explanation in first paragraph?
SeriousAndBenignInFirstPara: Serious/benign first para.?
NoSeriousBenignInFirstPara: No serious/benign first para.?
NumSerious: Num serious illnesses
NumBenign: Num benign explanations
NumGraveConcerns: Num grave concerns (e.g., fatal)
FracPageFirstSerious: Fraction page to first serious illness
FracPageFirstBenign: Fraction page to first benign explan.
NumSeriousInFirstPara: Num serious illness in first para.
NumBenignInFirstPara: Num benign explan. in first para.
NumNegMod: Num negative modifiers (e.g., don't have)
NumPosMod: Num positive modifiers (e.g., do have)
NumTestimonials: Num testimonials (e.g., I was scared)
UrlTrusted: Page from trusted source (e.g., medlineplus)?
TrustedDomain: Page from trusted domain (e.g., .edu)?
IsWebForum: Page from a Web forum?
HasURACVerification: Verified by www.urac.org?
HasHONVerification: Verified by www.healthonnet.org?
HasSeekMedicalAdvice: Recommends medical consult.?
ForHealthProfessionals: Content meant for health prof.?
LengthInWords: Num words
SizeInKB: Size in kilobytes (text only)
HasResources: Mentions external resources (e.g., doctor)?
Session Features (from concern to HUI query or session end)
NumQueries: Num queries
AvgQueryLength: Average query length (in tokens)
NumEscQueries: Num queries with escalations for concern
NumNonEscQueries: Num queries with benign explanation
NumURLs: Num (non- search engine result) pages
AvgDwellTime: Average dwell time on pages
TotalDwellTime: Total dwell time on pages
User Features (from sessions preceding the current session)
AvgConcernSearchesPerDay: Num concern queries per day
AvgMedicalSessionsPerDay: Num medical sessions per day
NumUniqueSymptoms: Num unique Merck symptoms
NumEscalations: Num previous queries for serious illnesses
NumResourceQueries: Num previous resource queries

Experimental Methodology

We constructed a training set containing pages accessed during sessions that contain queries on basic symptoms. Each page was tagged as appearing im-

mediately preceding an HUI query or not. The training set contained an equal number of pages adjacent to HUI queries and pages adjacent to session terminations (such that the accuracy of a marginal model that always predicted healthcare utilization intent was 50%), and used five-fold cross-validation.

Prediction Findings

We computed the accuracy of models built from features of the page; features of the page and features of the session; and features of the page, the session, and the current user. The best classification accuracy was obtained when all features classes were used (77.7%) (the *full model*). The inclusion of user information significantly increased accuracy over page and session features combined (68.9%) and page features alone (59.3%). All differences between models were significant at $p < .01$ using analyses of variance and post-hoc testing as appropriate across 100 runs. Figure 3 shows the receiver-operator characteristic (ROC) curves averaged across all runs.
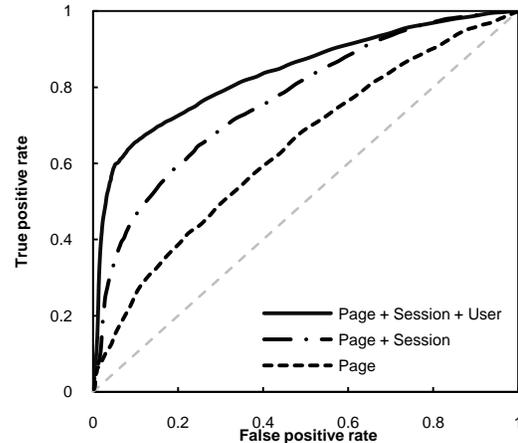

Figure 3. Receiver operator characteristic curves.

The curves in Figure 3 show that the performance improves as models that consider only page features are extended with session and user features. Search behavior during the session may provide insight into the nature of the search task, helping the model to classify more accurately. User features improve performance, perhaps since predispositions or search habits contribute significantly to HUI queries. Table 3 shows the five most important features in the full model, as well their classes and weights relative to the most predictive feature, *AvgDwellTime*.

The findings illustrate the importance of all feature classes in the prediction of whether the action following the page will be an HUI query or session end. Although the average dwell time on pages is the most important feature in our model, importance weights show that the other features are also valuable.

Table 3. Top features by evidential weight in full model.

Feature	Class	Weight
AvgDwellTime	Session	1.00
NumEscalations	User	0.71
HasResources	Page	0.60
NumResourceQueries	User	0.56
NumURLs	Session	0.47

Discussion and Implications

We have demonstrated significant differences in features of the pages, the sessions, and the users for sessions containing HUI queries versus sessions that terminate without them. We also showed that the search for information on specific treatment location centers or specific physicians are the most common HUI queries and demonstrated that HUI queries typically occur near the end of search sessions. In cases where the HUI query is not the terminal session query, it appears that users often refine their searches to polish the targeting of a local healthcare resource.

We built models to predict escalations from symptoms to searches on local healthcare resources and showed that we can achieve fair accuracy using features of adjacent pages. Classification accuracy improves significantly with the addition of session and user features, signifying the value of this information even with only a small number of fairly rudimentary features. Session and user features could be enhanced to aggregate page features across multiple pages and include features of search engine result pages.

Our findings offer insights about the goals of people turning to the Web for health concerns. The findings have design implications for features and services that can address potentially unnecessary and costly concerns associated with perusing medical Web content. Accurately predicting when users will pursue local medical resources may help search engines reduce unneeded engagement with health professionals, perhaps via user alerting mechanisms, ranking enhancements, or interface changes. Engagement with health professionals may be necessary when medical concerns are founded, and information such as the prior and posterior likelihood of concerning diagnoses can help guide users on their concerns and help them with healthcare resource utilization decisions.

We note that future systems might render session features unavailable for predictive models per user preferences regarding the privacy of their search data. There is promise in developing methods that allow searchers to opt-in to extended services that make ideal use of data that searchers are comfortable with sharing with a search service.[14] Other approaches to learning and using predictive models while enhancing privacy include the cloaking of personal data via aggregations of data from multiple users.

Conclusions

We presented results of a study on the escalation of Web searches on medical symptoms to queries pursuing in-world healthcare assistance as proxies for local healthcare resource utilization intent. We characterized aspects of resource usage and developed models to predict transitions to queries for external resources using features of the current Web page, the session, and aspects of the user's medical search history. Finally, we discussed the potential to harness inferences about the likelihood that searchers will pursue unnecessary healthcare resource utilization in the design of future services and interfaces.

References

1. Pew Internet and American Life Project. Online Health Search 2006. Accessed September 2008.
2. White, RW, Horvitz, E. Cyberchondria: Studies of the escalation of medical concerns in web search. ACM Trans on Inf Sys, 2009; 27(4).
3. Wagner TH et al. Does providing consumer health information affect self-reported medical utilization? Medical Care, 2001; 39: 836-847.
4. Baker, L et al. Use of the internet and e-mail for health care information. J Am Med Assoc, 2003; 289(18).
5. White, RW, Horvitz, E. Experiences with web search on medical concerns and self-diagnosis. In Proc. AMIA, 2009, 696-700.
6. White, RW, Horvitz, E. Predicting escalations of medical queries based on web page structure and content. In Proc. SIGIR, 2010, in press.
7. Kassirer, JP. Patients, physicians, and the internet. Health Affairs, 2000; 19(6): 115-123.
8. Murray, E et al. The impact of health information on the internet on health care and the physician-patient relationship. J. Med Inter Res, 2003; 5(3).
9. Nicholson, W et al. The association between women's health information use and health care visits. Women's Health Issues, 2005; 15: 240-248.
10. Eastin, MS, Guinsler, NM. Worried and wired: effects of health anxiety on information-seeking and health care utilization behaviors. Cyberpsychology and Behavior, 2006; 9(4): 494-498.
11. Bengeri, M, Pluye, P. Shortcomings of health-related information on the internet. Health Prom Int, 2003; 18(4): 381-387.
12. Cline, RJ, Haynes, KM. Consumer health information seeking on the Internet: the state of the art. Health Ed Res, 2001; 16(6): 671-692.
13. Eysenbach, G, Köhler, C. How do consumers search for and appraise health information on the world wide web? Brit Med J, 2002; 324: 573-577.
14. Krause, A, Horvitz, E. A utility-theoretic approach to privacy and personalization. In Proc. AAAI, 1181-1188.