Predicting Escalations of Medical Queries Based on Web Page Structure and Content

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

ABSTRACT
Logs of users' searches on Web health topics can exhibit signs of escalation of medical concerns, where initial queries about common symptoms are followed by queries about serious, rare illnesses. We present an effort to predict such escalations based on the structure and content of pages encountered during medical search sessions. We construct and then characterize the performance of classifiers that predict whether an escalation will occur after the access of a page. Our findings have implications for ranking algorithms and the design of search interfaces.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – *search process*, *selection process*

General Terms
Experimentation, Human Factors

Keywords
Medical search, cyberchondria

1. INTRODUCTION
Web search engines are used frequently by consumers to access health information [4]. Although such retrieved information can be invaluable, problems may arise when Web search is relied upon for diagnosis, where queries describing symptoms are input and rank and results are interpreted as diagnostic conclusions. Prior research identified search and browsing sessions where queries on symptoms associated largely with common, benign explanations can lead to searches on more serious, rare ailments [7]. Such escalations may be based in several factors, including the relatively large quantity of Web content describing serious illnesses versus benign explanations, and the potential use of ranking algorithms based on clickthrough data. Presenting people with troubling health scenarios without information on the probabilities of these outcomes, or on the typically greater likelihoods of more common explanations, can heighten concerns inappropriately [6].

Previous studies have explored the effects of factors such as page design and content on the perceptions of health information seekers [3][5] and on actions such as self-diagnosis and self-treatment [1], and have showed that, while 80% of American adults have searched for healthcare information online, 75% refrain from verifying key quality indicators such as source validity and source creation date [4]. We focus here on predicting escalations in queries based on the structure and content of Web pages. We specifically build classifiers to predict whether an escalation will occur immediately following a page visit. Accurate predictions of escalations in medical concerns given page features allows search systems to flag specific pages as potential sources of inappropriate anxiety or down-weight such pages in their result ranking.

2. IDENTIFYING QUERY ESCALATIONS
We studied anonymized logs of URLs visited by users who opted in to provide data through a widely distributed browser toolbar. Log entries include a user identifier, a timestamp for each page view, and the URL of the page visited. Intranet and secure (https) URL visits were excluded at source. Only entries generated in the English speaking United States locale were included. From these logs, we mined many thousands of search sessions. Sessions start with a search engine query followed by a click on a search engine result. We consider sessions to be a sequence of time-ordered URLs terminating after 30 minutes of user inactivity (as in [7]).

We define query escalations as observed increases in the medical severity of search terms used within single health-related search sessions. For the purpose of the study, we examined the query terminology for sessions containing queries on six common symptoms: *headache, chest pain, muscle twitches, abdominal pain, nausea,* and *dizziness*. We identified cases where users transition from a query about one of these basic symptoms to queries on a related, but serious illness within a search session. We focus on two outcomes following a query about a basic symptom:

1. *Next-query escalation*: Following visits to pages after a query, an escalation occurs on the *next* follow-up query.
2. *Any-query escalation*: Following visits to pages after a query, an escalation occurs on *any* follow-up query within a session.

Lists of terms for expressing the six basic symptoms and terms for expressing serious illnesses for each of the symptoms were constructed through review of the medical literature as well as the browsing of portions of the search logs by the authors, one of whom has received medical training during an PhD/MD program (EH). Lookup into these lists was used to identify candidate sessions and escalations in the queries present in session logs.

Predicting escalations requires identifying non-escalations as well as escalations. Non-escalations are defined in several ways:

1. Follow-up query is benign as explicitly defined.
2. Follow-up query not associated with defined escalation or non-escalation (hereafter called an *undefined non-escalation*).
3. Session terminates without escalation.

Figure 1 illustrates these definitions graphically, with q as queries, boxes as pages, and P as the page from which features would be generated in each case. In our study, we experiment with the prediction accuracy for each of these definitions of non-escalations.

3. PREDICTING ESCALATIONS
We constructed classifiers to predict whether a user's next query following a page visit contains an escalation or a non-escalation based on features of that page. We built the models using logistic regression, considering the outcome for each identified session and a set of over 40 features representing attributes of the structure and content of pages. Space limitations preclude detailing the entire feature list. However, in summary the features used include:
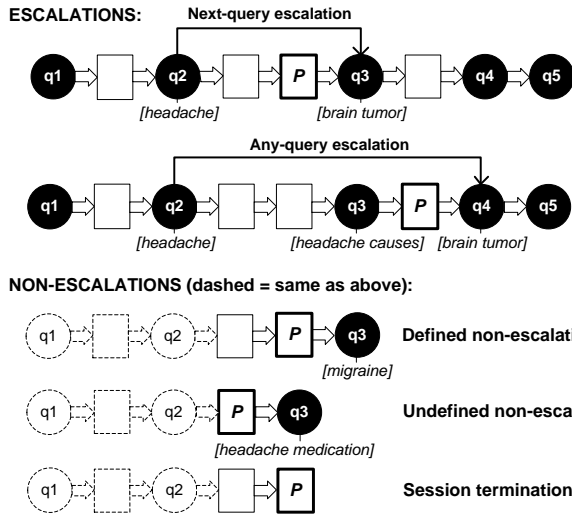
Figure 1. Escalations and non-escalations in search sessions.

structural features (e.g., serious illness precedes benign explanation, number of words between serious illness and benign explanation, modifiers such as "unlikely to be" appear near serious illnesses), title and URL features (e.g., title has serious illness, is forum URL), first-person testimonials (e.g., page has phrases such as "told me", "I felt", and "I am worried"), page reliability features (e.g., page is externally verified by healthonnet.org, recommends consulting physician), commercial intent features (e.g., page has advertisements, number of advertisement blocks), and general page features (e.g., length in words, total kilobytes).

We used a set of six thousand Web pages sampled randomly from our data. In initial analyses, we separately probed the predictive value of single features. For example, we considered the influence on escalation of the order in which serious illnesses versus benign explanations appeared on the pages viewed (for pages containing both serious illnesses and benign explanations). Table 1 shows the boost in escalations associated with viewing pages that relay information about serious disorders before discussing benign explanations as compared with viewing pages with the reverse ordering. One explanation for this phenomenon is that users may not read the full page (as supported by eye-tracking research [2]).

Each page is associated with an escalation or non-escalation in the query immediately following it. We selected an equal number of escalations and non-escalations (such that the accuracy of a marginal model that predicted escalation was 50%), and used five-fold cross-validation. Prediction accuracy varied with definitions of escalation and non-escalation. The highest accuracy was obtained for next-query escalations and defined non-escalations (73.4%), followed by next query escalation where non-escalation is session end (70.7%), and next-query escalation where the escalation is undefined (68.8%), and finally any-query escalation (65.5%). Performance differences were statistically significant (at $p < .01$) between all methods across 100 runs. Figure 2 shows the receiver-operator characteristic curves for one of the runs.

Our findings show that we can generate potentially valuable predictions about escalations using page features, and that accuracy varies with the definition of escalations and non-escalations. Best performance is obtained when non-escalations come from a defined set of benign explanations with no queries between concern and escalation. The most predictive features for all models are:

Table 1. Percentage of escalations versus non-escalations given order of serious and benign conditions on the preceding page.

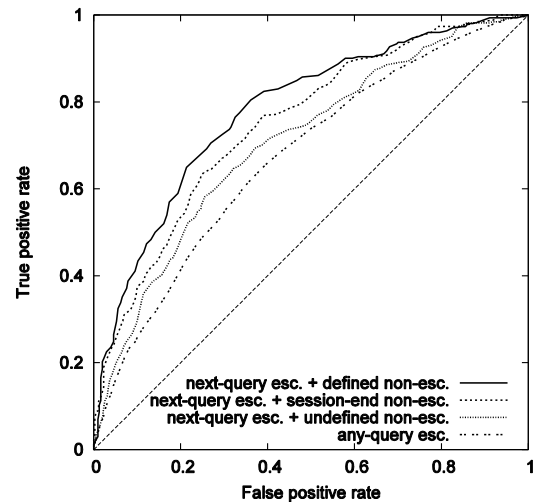| Query outcome	Order of presentation	
	Serious first	**Benign first**
Escalation	68.6%	33.4%
Non-escalation	31.4%	66.6%

Figure 2. Receiver-operator characteristic curves.

serious illness precedes benign explanation in page, serious illness vs. benign explanation appears in page title or near beginning of page, page from Web forum, and page has external verification.

4. SUMMARY

We studied the prediction of escalations of queries for health information on the Web based on features of accessed pages. The results have implications for the design of search engines that may use escalation likelihood as a ranking feature or present this information to users. Future directions include the leveraging of the richer rhetorical structure of page content to introduce additional structural features. We also hope to investigate the predisposition to escalate based on features of users, examine cumulative effects of multiple session pages, and study the influence of contents displayed on search results pages.

REFERENCES

[1] Bengeri M. & Pluye P. (2003). Shortcomings of health-related information on the internet. *Health Prom Int.*, 18(4): 381-387.

[2] Buscher, G., Cutrell, E. & Morris, M.R. (2009). What do you see when you're surfing? Using eye tracking to predict salient regions of web pages. *SIGCHI*, 21-30.

[3] Eysenbach, G. & Köhler, C. (2002). How do consumers search for and appraise health information on the world wide web? *British Medical Journal* 324: 573-577.

[4] Fox, S. (2006). Online health search 2006. Pew Internet and American Life Project. Accessed January 2009.

[5] Sillence, E., Briggs, P., Fishwick, L. & Harris, P. (2004). Trust and mistrust of online health sites. *SIGCHI*, 663-670.

[6] Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: heuristics and biases. *Science*, 185(4157): 1124-1131.

[7] White, R.W. & Horvitz, E. (2009). Cyberchondria: Studies of the escalation of medical concerns in web search. *TOIS*, 23(4).