Anxiety and Information Seeking: Evidence From Large-Scale Mouse Tracking

Brit Youngmann
Microsoft Research
Herzliya, Israel
t-bryoun@microsoft.com

Elad Yom-Tov
Microsoft Research
Herzliya, Israel
eladyt@microsoft.com

ABSTRACT

People seeking information through search engines are assumed to behave similarly, regardless of the topic which they are searching. Here we use mouse tracking, which is correlated with gaze, to show that the information seeking patterns of people differ dramatically depending on their level of anxiety at the time of the search.

We investigate the behavior of people during searches for medical symptoms, ranging from benign indications, where users are not usually anxious, to ones which could harbinger life-threatening conditions, where extreme anxiety is expected. We show that for the latter, 90% of people never saw more than the top 67% of the screen, compared to over 95% scanned by people seeking information on benign symptoms, even though relevant documents are similarly distributed in the results pages to these queries. Based on this observation, we develop a model which can predict the level of anxiety experienced by a user, using attributes derived from mouse tracking data and other user interactions. The model achieves Kendall's Tau of 0.48 with the medical severity of the symptoms searched.

We show the importance of using information about the users' level of anxiety as predicted by the model, when measuring search engine performance. Our results prove that ignoring this information can lead to significant over-estimation of performance. Additionally, we show the utility of the model in three special instances: where multiple symptoms are searched concurrently; where the searcher has an underlying medical condition; and when users seek information on ways to commit suicide. In the latter, our results demonstrate the importance of help-line notices, and emphasize the need to measure the effective number of results seen by the user.

Our results indicate that measures of relevance which use anxiety information can lead to more accurate understanding of the quality of search results, especially when delivering potentially life-saving information to users.

CCS CONCEPTS

• **Information systems** **Search interfaces**; **Relevance assessment**; • **Applied computing** *Consumer health*; *Health informatics*;

KEYWORDS

Mouse tracking, Relevance, Health, Medicine

ACM Reference Format:
Brit Youngmann and Elad Yom-Tov. 2018. Anxiety and Information Seeking: Evidence From Large-Scale Mouse Tracking . In *WWW 2018: The 2018 Web Conference, April 23–27, 2018, Lyon, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3178876.3186156

1 INTRODUCTION

Web search engines allow users to search and retrieve relevant information from billions of Web pages. Typically, a user issues a search query and the engine returns a list of results, ranked according to the documents' *relevance*. The relevance of the results is inferred from a number of factors [20], including: (i) how well a document matches the user's query; (ii) the document's reputation, and (iii) implicit feedback inferred from the users' behaviors for that query. Records of the latter include clicks made by the user and more recently, data on the movements of the users' mouse position on the search engine results page (SERP), which provide an implicit signal for the relevance of results [23].

Mouse (or cursor) tracking [24] is the use of software to collect the positions of the users' mouse cursors on the computer or browser page. These data are gathered to obtain richer information on the interaction between the user and a computer or a website, typically to improve the design of an interface [33], to measure relevance [20] or, more recently, to estimate search satisfaction, attention and interest [5, 26, 29]. Eye gaze, that is, what a user is looking at, has been shown to be correlated with the position of the cursor [14]. Thus, mouse tracking has been used as a proxy for eye tracking in large-scale experimentation for measurement of user attention in web search, often to collect information when users do not explicitly click on web pages and instead only perform pointing actions [22]. Outside of search engine design and human-computer interfacing, eye tracking has been used as a powerful tool in experimental psychology, as it provides data which reflects the cognitive and psychological states of individuals [14].

Most people feel anxious from time to time. Barlow [9] defined anxiety as "a future-oriented mood state in which one is not ready or prepared to attempt to cope with upcoming negative events". *State anxiety* is a widespread reaction to a stressful situation [39], and can be defined as fear, nervousness, discomfort, and the arousal of the autonomic nervous system induced temporarily by situations perceived as dangerous, i.e., how a person is feeling at the time of a perceived threat [38]. State anxiety often impairs physical and psychological function. Common symptoms of state anxiety which are manifested in vision are tunnel vision, blurred vision and double vision [10]. In this paper, the terms anxiety and state anxiety are

WWW 2018, April 23–27, 2018, Lyon, France

ACM ISBN 978-1-4503-5639-8/18/04..
https://doi.org/10.1145/3178876.3186156

used interchangeably, although in all cases we refer to state anxiety. We note that anxiety should not be confused with *anxiety disorder*, a mental disorder characterized by significant feelings of anxiety, which we do not investigate in this paper.

Here we use mouse tracking data to understand how people's interaction with search engines changes when they experience anxiety owing to the nature of the information they are seeking. Our hypothesis is that the more a user is anxious, because she is experiencing a stressful situation, as when she is having severe chest pains, the less she tends to explore the SERP presented to her, viewing only the top-ranked results. If true, this has important implications to search relevance, as we show in our Experiments below.

Consider the following illustrative example. Figures 1 (a) and (b) show heat-maps [20] describing the mouse movements of 441 people who made queries about chest pain and 472 queries about constipation, respectively (details on data collection are given in Section 3). In these heat maps, redder regions correspond to areas that users spent more time looking at, according to mouse tracking. One can see that users who asked questions about more stressful symptoms (chest pains) tended to look only at the top ranked answers, while completely ignoring all other results. In contrast, users who asked about less stressful conditions (constipation), tended to explore more of the results shown to them.

We focus on medical symptoms search, since prior work shows that the more severe the medical symptom experienced by a person, the more anxious she will be [11, 32]. We show that the anxiety level of users can be inferred from the topic of their queries (i.e, the symptom mentioned), and that user interaction with the SERP can predict the level of anxiety. We examine the importance of using information on the users' anxiety level when measuring search engine performance, and find that ignoring it can lead to significant overestimation of performance. Commonly used evaluation metrics for information retrieval ignore such information, and may thus lead to incorrect conclusions on the effectiveness of search results.

Additionally, we demonstrate the use of the model in three special instances: where multiple symptoms are searched concurrently; where the searcher has an underlying medical condition; and when users seek information on ways to commit suicide. In the latter, our results show the importance of help-line notices, and highlight the critical importance of measuring the effective number of results seen by the user.

To the extent of our knowledge, we are the first to use mouse tracking data to identify the underlying emotional state of users (specifically, anxiety) prior to the search action and independent of the content presented, and believe that our suggested methodology opens up an opportunity to a wide range of research concerning psychological aspects of the users vis-à-vis the interaction with search systems.

2 RELATED WORK

2.1 Medical Symptom Search

The Web is the first stop for a vast majority of Internet users who experience a medical symptom and are seeking information about it [43, 45]. Indeed, 80% of American adults have searched for medical information online [18]. The information obtained from medical searches can influence users' concerns, their decisions about when to engage a physician, and their overall approach to their health condition [42, 44, 45].

White and Horvitz [42] explored the relationship between the types of medical symptoms searched and the time taken to visit a medical facility. Their results indicated a strong dependence of the time between which a user queried for a medical symptom and the time she first arrived to a medical facility to treat the symptom she queried for. These time differences were significantly lower for symptoms which may be more worrying to users, such as chest pain, compared to more benign symptoms such as constipation or nausea. Thus, interaction with information on the Web reflects the anxiety level of a user and suggests itself as a method to infer it.

2.2 User Behavior Analysis

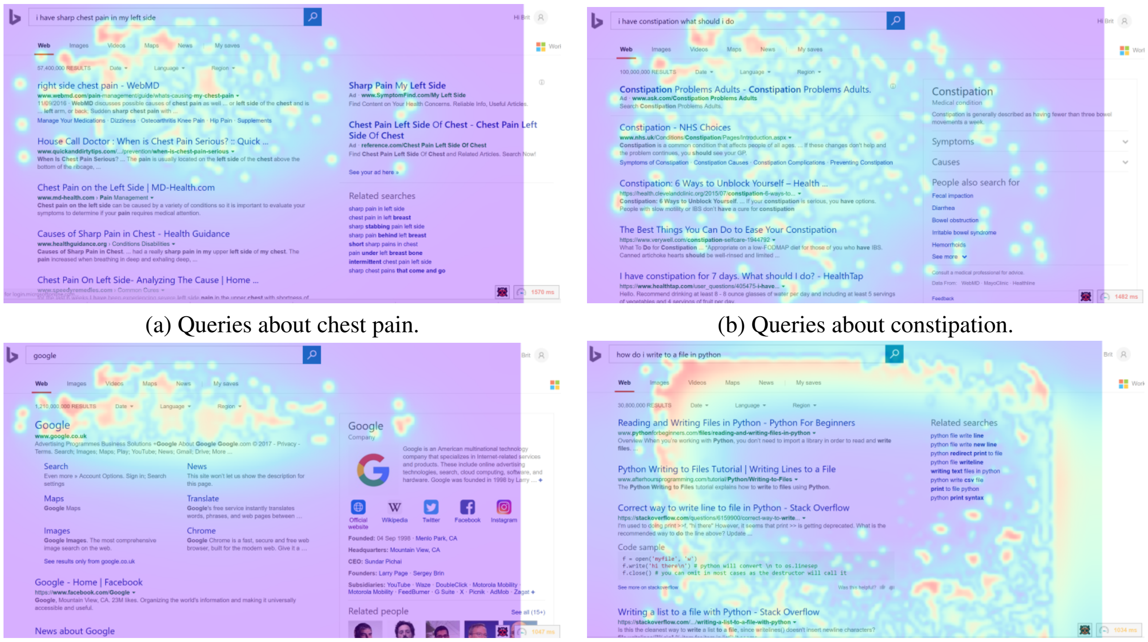
Understanding how users interact with the SERP is a fundamental question in information retrieval, bearing on relevance evaluation, search quality, and interface design [3, 24, 29].

Result click-through statistics and dwell times on clicked results have significant value for inferring the relevance of search results [20]. However, the interpretation of such signals can vary substantially for different search queries and users, and it provides little information on what parts of the SERP the user examined.

Previous work has suggested the use of cursor movements to understand user behavior (e.g., [3, 7, 17]), as a cheap alternative to eye-tracking. The relationship between cursor and gaze was studied in depth [34], and cursor position was shown to be correlated with gaze when users performed clicks or pointing actions in search contexts. Therefore, these data have been successfully used to measure user attention in web search [22]. Specifically, mouse tracking data was used to infer content salience (e.g., [29]), improve ranking by estimating the relevance of results (e.g., [20, 24]), and dynamically estimate the result that searchers will request next (e.g., [16]). In contrast, eye and cursor movement are poorly coordinated during cursor inactivity [14]. This limits the utility of such data as an attention measurement tool in content reading tasks (e.g., news reading).

These works, while resembling ours in attempting to use mouse tracking signals to understand user behavior, differ in that in our work we attempt to quantify the users' level of anxiety prior to their access of the search engine, and, more generally, estimate their emotional state. Furthermore, their estimation of document relevance commonly ignores the text of the user's queries. In our work, these queries provide essential information on the users' emotional state, and, as we demonstrate, their emotional states directly affect the interaction with the SERP.

Eliciting the feelings of users while they seek information on the web is of high value, as it can improve search and user experience [6]. Previous work suggested analyzing cursor movement to estimate search satisfaction [15, 26], infer the interest of users in online content [5, 6, 30], or deduce searcher attention [29]. Different from relevance, interest or satisfaction prediction researches, in which *user behavior and emotional state in response to the content presented are inferred* and her underlying emotional state prior to the search is not considered, in this work we aim to infer how this emotional state affects her behavior. We take here the initial steps towards a better understanding of how *the user's mood*, in particular anxiety or stress, affects her interaction with the SERP.


i have sharp chest pain in my left side
i have constipation what should i do
google
how do i write to a file in python

(a) Queries about chest pain.

(b) Queries about constipation.

(c) Navigational queries.

(d) Informational queries.

Figure 1: Saliency maps for mouse cursor movements. Redder shades correspond to longer dwell times of the cursor, indicating higher interest of the user. [Best viewed in color]

2.3 Query Classifications

A widely-used taxonomy of web searches [12] distinguishes between three main types of user queries: (i) navigational queries, where the goal of the user is to reach a particular site that she has in mind, (ii) informational queries, where the goal of the user is to find information assumed to be available on the web, and (iii) transactional queries, where the intent is to perform some web-mediated activity. Guo and Agichtein [19] showed that a user's attention, as evident in the movements of the mouse, can be used to distinguish navigational from informational and transactional queries: the movements of the mouse are mostly confined to the top of the screen in the case of navigational queries, whereas in the case of informational/transactional queries such movements encompass the entire screen.

Cartright et al.[13] distinguished between two types of exploratory health search queries: (i) hypothesis-directed queries, where the goal of the user is to find content on one or more illnesses, including risk factors, treatments and therapies, and (ii) evidence-directed queries, where the intent is to understand the relevance of a set of observed symptoms. These two types of health search queries differ in that in the latter case, the user is not aware of her underlying health condition, and hence searches for information on the Internet.

3 EXPERIMENTAL METHODOLOGY

Our goal is to predict the level of anxiety or stress of a user, according to her engagement with the SERP. Toward that end, we next briefly present the methodology used in our experiments. We start by explaining how the mouse tracking data was extracted and used, then, explain how we estimated the level of anxiety. Finally, we explain the model and methods used in our experiments.

3.1 Data

We extracted queries from the USA to the Bing search engine which mentioned some medical symptom, between December 1st, 2016 and May 31st, 2017. We have collected over 22K user queries, asked on different medical symptoms (see Table 2). According to the users' unique identifiers, the queries were asked by approximately 21K different users. The medical symptoms considered in our experiments include the ones used in [42], and are common medical symptoms according to Wikipedia [2]. We excluded symptoms with fewer than 150 queries, thus retaining queries on 23 medical symptoms with 150 or more queries on each one of them.

For each user query we extracted the list of displayed and clicked results, and the mouse tracking data collected. To ensure the user asking the query is the one having the medical symptom, queries were then filtered to include those whose text contained a self description using one of the three phrases “I have”, “I'm having" or “I am having”. Furthermore, to ignore extreme cases that may affect the results, unless otherwise stated, we removed all queries that contain more than one medical symptom (e.g., “Why do I have low back pain and nausea") or implied an underlying health condition which could be the source of additional anxiety (besides the searched medical symptom). For example, a pregnant woman asking about a headache may experience anxiety related to her pregnancy in addition to that associated with the symptom [21].

Importantly, all procedures performed in this study were approved by the Institutional Review Board of the institution.

It is to be observed that some symptoms appear more frequently in user search queries than others (e.g., cough is a more common symptom than chest pains). However, for the sake of the learning-to-rank task that we employed, the (unequal) number of queries on each symptom is irrelevant, since we consider all pair-wise possible comparisons during the training process.

Search engines nowadays often provide an additional text box within the SERP that contains summarized information (referred as a quick answer), and/or a list of query suggestions. The sizes of these boxes may be larger than a single organic search result. Our results suggest that there is no correlation between the symptom queried and the probability such a box would be presented. Thus, we ignore the difference in quick answer sizes compared to those of organic results and note that rank 1 results may in fact be quick answer boxes.

The queries extracted in this study are all evidence-based queries [13]. Therefore, we assume that at the time of issuing the query the user is unaware of her underlying medical condition (and consequently may experience state anxiety or stress), leading us to assign the same rank to all queries mentioning the same symptom. Another underlying assumption is that the user asking the query is experiencing the symptom themselves. Previous work suggests that this is the case in the vast majority of medical symptom search queries [46].

We noted that the potential errors in the analyses include: (i) The user is in fact not experiencing the mentioned symptom, e.g., "I feel weak but I don't have a fever", and (ii) The user does not necessarily suffer from the medical symptom at the time of asking the query, e.g., "what does it mean when I have had a headache for 2 days?" We further performed explicit filtering to reduce the likelihood of these errors in our data, removing all queries that contained the string "don't have".

3.2 Mouse Tracking

Mouse tracking data consist of a list of time-stamped horizontal and vertical coordinates of the mouse cursor location during the interactions of the user following a search query. We represent these data by extracting summarizing features such as minimal or maximal points on the screen. Table 1 provides the full list of all extracted features, including features extracted from the displayed and clicked results. These features were chosen so as to quantify the amount of the user interaction with the SERP, and the extent of the SERP with which the user interacted. For example, we counted the number of times the user scrolled up and down the SERP by counting the number of local vertical (y) minima on the page.

Another example for a summarizing feature is the number of clicked results out of the number of displayed results. This feature provides information on how many results the user looked at (and found relevant). However, as this feature ignores the positions of the clicked results, we also counted the number of clicked results below a certain index (i.e., the 1-st, 3-rd or 5-th indexes). The addition of clicks below other positions was empirically found to be superfluous. Other features such as the velocity, acceleration, and jerk of the mouse movement were considered, but were found not to add information and were thus excluded in the following analysis.

3.3 Estimating the Medical Severity Rank of Symptoms

Our analysis of the symptoms is performed against a measure of their severity, which we refer to as the *Medical Severity Rank* (MSR) of the symptoms. The levels of MSR were defined by medical experts and, as we show below, are highly correlated with a previously-proposed measure of symptom severity and with the ranking provided by non-experts.

We recruited 3 medical professionals (two medical doctors and one registered nurse) from the website oDesk.com to rank the set of symptoms. The professionals were asked to assume that someone is experiencing each of the symptoms (separately) and rank them on a Likert scale of 1 to 10 on how urgently this person should seek medical attention, where 10 indicates the symptom is not worrying at all and the person can disregard it, and 1 means that she should immediately go to the nearest hospital or call an ambulance. Note that the experts received only the set of symptoms, not user queries.

The 3 scores were averaged, and this was the score used for each symptom. The average Spearman's ρ correlation among the professionals was 0.60 ($P < 0.05$ in all pair-wise comparisons), suggesting a strong agreement on the severity level of the symptoms.

Several symptoms were given an average score that was of similar value. Assuming that similar scores represent the same level of severity (i.e., rank), we grouped together symptoms with scores closer than 0.33 of each other. This resulted in seven distinct ranks[1], where 7 is given to the least serious medical conditions and 1 to highly severe medical conditions, as depicted in Table 2.

In contrast to medical professionals, laypeople (who are the majority of searchers online) may not comprehend the severity of their symptoms. Therefore, to estimate how MSR correlates with laypeople's understanding of symptom severity, we recruited 15 student volunteers to assess the severity of the considered symptoms and to estimate the level of anxiety they think they would experience if they were suffering from one of the symptoms. The participants were asked to assume they are experiencing each of the symptoms (separately) and rank them on a Likert scale of 1 to 10 on how urgently they would seek a medical attention (here again, 1 means that the they would immediately go to the nearest hospital or call an ambulance and 10 that they would disregard the symptom). Additionally, they were asked to estimate on a Likert scale of 1 to 10 how anxious they think they would feel in each case. The average Spearman's ρ correlations among the participants were 0.58 and 0.55, the urgency and anxiety ranks, respectively ($P < 0.05$ in all pair-wise comparisons). Here again, we computed the average scores among all participants, resulting in two aggregated ranking lists.

We compared ranking given by the medical professionals with the ranking given by the volunteers. The Spearman's ρ between MSR and the urgency rank is 0.67 ($P < 0.05$), and between MSR and the reported anxiety rank is 0.61 ($P < 0.05$), suggesting that the MSRs are strongly correlated with laypeople's point of view. Importantly, this experiment demonstrates the connection between the severity level of a medical symptom (as estimated by non experts) and the level of anxiety it potentially causes. Not surprisingly, we see that the more severe the symptom is according to the non-experts opinions, the more anxiety is assumed. The Spearman's ρ between the urgency and assumed anxiety ranks is 0.72 ($P < 0.05$).

Each of these rankings is imperfect: Medical professionals know more than laypeople and assign slightly different importance to symptoms. On the other hand, the laypeople who volunteered to

[1] Since we are interested in the ranks and not in the absolute scores of symptoms, we consider the obtained scores on an ordinal scale, i.e., disregarding the differences between scores.

Features
Max x,y point
Mean x, y point
Min x, y point
Variance of x,y
Number of local y minimums
Duration of session
Total mouse distance
Rank of deepest clicked result
Fraction of displayed results that were clicked
Number of clicks below the 1-st, 3-rd or 5-th index
Percentage of the screen seen (width or height)

Table 1: Features from users' interactions with the SERP through both mouse movements and clicks. x and y are the horizontal and vertical screen coordinates of the cursor, resp.

provide these scores did not experience most of the symptoms themselves, and could only assume how anxious they would feel. We chose to use the MSR as given by medical professionals herein, and note the high correlation between it and the scores of laypeople. However, the correlation with the actual anxiety of users likely lies between that of laypeople and professionals.

Previous research has shown that the more serious the medical symptom is, the more stressed and anxious the user will be [11, 32], whether they are searching for themselves (as is the majority of cases [46]) or for close family members.

We correlated the resulting MSR of symptoms with the results given by White and Horvitz [42], who measured the time between when a user issued a symptom search query until the time when evidence for healthcare utilization (EHU) exists. EHUs in that study were evidence that the user is near a medical facility. In their work, the authors considered only a portion of the symptoms considered in this study (13 symptoms out of the 23 symptoms examined here). The Spearman's ρ between MSR and EHU is 0.51 ($P < 0.05$), suggesting that MSRs and EHUs are similar, though not identical, measures of symptom severity. Interestingly, excluding headache as one of the symptoms, the Spearman's ρ is 0.65 ($P < 0.05$), as according to the medical experts the MSR of headache is 5, while its EHU is 2. That is, the medical experts did not consider headache as a severe medical symptom, while the evidence suggested that people having a headache do not wait long until they first visit a medical care center. This may be because people query for headache only when they experience a severe manifestation of the symptom, and thus are likely to seek medical treatment, whereas the experts considered headaches in general.

Aside from headaches, MSRs and EHUs have a very high agreement on low ranked symptoms, i.e., symptoms that the medical experts ranked with a MSR of 4 or less: Here Spearman's ρ is 0.89 ($P < 0.05$). Thus, when symptoms are generally worrying, both their severity and the time for treatment seeking are highly correlated.

3.4 Ranking the Symptoms

MSRs were calculated for 23 symptoms. However, the gamut of symptoms is far larger. Therefore, to be able to estimate the MSR of symptoms outside the list of 23 symptoms, we trained a ranker to predict the MSR from user interactions with the SERP. The ranker is trained in the Learning-To-Rank (LTR) framework.

Symptoms	MSR	# of queries
Constipation, Nasal congestion	7	1031
Joint pain, Cough, Sore throat,Fatigue	6	3612
Headache, Earache, Diarrhea, Hip pain, Knee pain, Neck pain	5	7425
Fever, Back pain, Nausea, Rash Swollen feet	4	7704
Dizziness, Vertigo	3	1342
Palpitation, Difficulty swallowing	2	504
Chest pain, Shortness of breath	1	908

Table 2: Symptoms, Medical Symptom Ranking (MSR) and the number of queries at each MSR contained in the dataset.

LTR solves a ranking problem on a list of items. The aim of LTR is to learn an optimal ordering of those items, while minimizing the number of inversions in ranking. As such, LTR is concerned with the relative scores of items, not their absolute scores. In our setting, we do not care about the exact amount of anxiety, as we aim to learn what kind of behaviors imply a more stressful situation than others.

In our experiments, we used the SVM^{rank} [27], a highly efficient LTR model based on an SVM model with a polynomial kernel of rank 2. We note that a dedicated solution for ranking from partial and biased information feedback (e.g., clicks) was recently proposed [28]. However, this implementation only supports binary labels on the training data (i.e., marking documents as relevant or irrelevant), therefore, it is not suitable for our settings.

Attributes of the examples used by the ranker are only based on the interactions of the users with the SERP, as described above.

The input of the SVM^{rank} model are the feature vectors corresponding to each user query, and its output are real numbers in [-1,1]. From these scores, the ranking can be recovered via sorting. In our case, since we wish to focus on the 7 discrete levels of MSR, we further cluster the outputs of the ranker into 7 clusters using k-means clustering on the output values of the ranker. The clusters are sorted according to the average value of the ranker outputs of examples in each of them, and each example is given the rank (between 1 and 7) of the cluster to which it was assigned. To maintain a balanced number of samples in the clustering, we randomly selected an equal number of samples from each MSR value.

3.5 Search Engine Performance

Different evaluation measures have been suggested to assess the quality of search engine results [35]. Arguably, the most common one is precision, which counts the number of relevant documents returned in response to a query. Precision is commonly evaluated for the top k documents retrieved, when it is known as precision at k ($P@k$). Precision does not take into account the position of the relevant documents among the top k results. To overcome this, Normalized Discounted cumulative gain (NDCG)[25] was suggested. NDCG measures the usefulness of a document based on its position in the result list, with the gain of each result discounted at lower ranks.

For assessing Bing performance in our experimental study, we pick a random sample of 65 symptom queries. We labeled the returned documents for their relevance by asking 5 crowdsourced workers from the CrowdFlower website (https://www.crowdflower.com) to rank the relevance of each of the top 10 pages displayed in response to the 65 queries by Bing. Users were asked to rate how well the page helps understand the nature of the problem or the solution to it on a 3-point scale, where a score of 0 means the document is not relevant at all and 2 implies a highly relevant result. We used the average score for each page to compute the NDCG, and when computing precision assumed that a page was relevant if its average score was equal to 2.

4 RESULTS

In this section we provide results of our efforts to validate the effectiveness of the proposed model. We then show the importance of estimating the fraction of the screen that users saw when evaluating the performance of search engines. Finally, we apply the model to learn the level of anxiety associated with three specific scenarios (using the technique described in Section 3.4).

4.1 Predicting Anxiety from User Interactions

First, we evaluate how the learning-to-rank model is able to capture the severity of symptoms (and hence, the likely anxiety level of users) from their interactions with the SERP. We quantify this using Kendall's tau rank correlation coefficient and Spearman's rank correlation coefficient. The performance of the model was evaluated using 10-fold cross validation, where each sample represents a feature vector, which in turn, represents a single user session (note that the feature vector does not contain the text of the query). We report the average score across all runs.

Using the trained model, we ranked the queries according to user interactions and measured the correlation of the predicted level of anxiety with the level associated with the text of the query (the MSRs ranks). We found that the average Kendall τ was 0.48, and the average Spearman $\rho = 0.40$ ($p < 0.05$ in all cases). Interestingly, when considering examples whose ranks are far apart, i.e., ranks with a gap larger than 2, the average values were $\tau = 0.55$ and $\rho = 0.53$.

Thus, user interaction with the SERP can explain a large portion of the level of anxiety one would predict that the user is experiencing, according to the text of the query. This experiment provides empirical evidence that the interaction with the SERP is affected by the user's emotional state, specifically, her level of anxiety.

It is possible that the observed correlations are the consequence of the search engine response to different queries. Namely, high-anxiety users interact with the top-ranked results more, as they are more relevant, while users who are less anxious receive a SERP with less-relevant results and therefore have to browse deeper into the SERP before their information need is met. To refute this, we computed the NDCG of the sampled queries and their crowd-labeled displayed results (as describe in section 3.5). The Pearson correlation between the NDCG scores and the MSRs was $r^2 = 0.03$ ($p > 0.2$).

Thus, the distribution of relevant results within the SERP cannot explain the ability to predict MSR from mouse movements. Moreover, as we discuss below, these movements indicate that anxious users did not even see the low-ranked results, and could not therefore determine whether the low-ranked results are relevant to their needs.

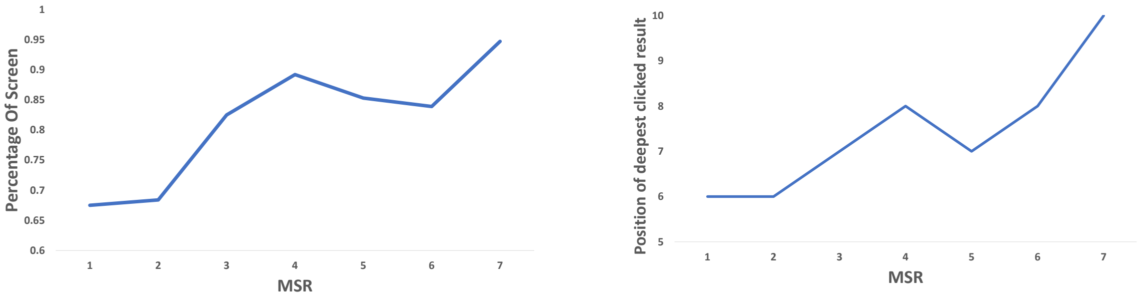
To demonstrate the association between the behavior of users and the associated MSR rank of their queries, we focus on two features extracted from the users' recorded sessions. Figures 2 show the dependence of two of the attributes used in the model: (1) The percentage of the salient screen (i.e., the portion of the screen the user interact with, based on the mouse tracking data) and (2) the position of deepest clicked result (where position 1 is the result at the top of the SERP) on the rank of the symptom, as given by the medical annotators. The horizontal axis represents the MSR of the symptoms (which serve as the ground truth), and the vertical axis represents the 90-th percentile value obtained for all samples in a given rank. For example, Figure 2 (a) shows that 90% of the users who asked a query associated with the highest severity saw only 67% or less of the SERP, compared to ones asked a query associated with the lowest severity (rank 7), that saw more than 95% of the SERP. These figures show that in lower ranked queries, i.e., those where the level of anxiety is higher, the user focuses on a smaller portion of the SERP and tends to click on higher ranked results. However, the correlation between the two attributes and the level of anxiety as implied by the text of the queries is not perfect: A model trained using only these two features provides inferior results, with an average Kendall's τ of 0.35 and an average Spearman's $\rho = 0.28$ ($p < 0.05$).

Table 3 depicts the average Kendall's τ achieved using single features, sorted (from top to bottom) by the average Kendall's τ score. For example, the first row shows that using only the fraction of the vertical salient screen yields Kendall's τ of 0.18. As the table shows, the most important features are the percentage of the salient screen and the closely correlated position of the deepest click.

Navigational and Informational Queries. The above analysis demonstrated that the more serious the symptom about which the user asks, the less of the SERP she interacts with. Another area where we expect a difference in the interaction of a user with the SERP is when one compares the interaction with the SERP in the case of informational query, compared to when a user issues navigational query. In the first, we expect users to interact with a large portion of the SERP (as in a less severe symptom) whereas in the latter, we expect users to interact with a small part of the screen (typically the first result). Thus, here we consider 1000 navigational queries and 2000 informational queries, and apply the model constructed above to them.

The navigational queries included were identified by extracting the list of the most popular Internet websites, as listed in the relevant Wikipedia page (https://en.wikipedia.org/wiki/List_of_most_popular_websites) and extracted queries which mention one of these sites (e.g., the search query "Google"). For informational queries, we extracted data on queries containing the text "How do I" or "How to", but do not contain any of the symptoms listed above.

Figures 1 (c) and (d) show heat-maps describing the mouse movements from the navigational and informational queries, respectively. Comparing these figures to Figures 1 (a) and (b), we observe that queries for a severe symptom (chest pain) are very similar to navigational queries, whereas queries for a benign symptom (constipation) are more similar to informational queries.



(a) Fraction of the vertical salient screen as a function of MSR. (b) Position of deepest clicked result as a function of MSR.

Figure 2: The connections between two features used and the MSRs.

We applied the model to the interaction data. As expected, 92% of the navigational queries were predicted to be of rank 1, with the remaining queries predicted to be of rank 2. Contrarily, 78% of the informational queries were predicted to be of rank 7, 16% of rank 6 and 6% of rank 5. The uncertainty concerning the informational queries stems from the large variance between different informational queries, i.e., each query is asking on a different topic.

Here again we considered the possibility that this effect was because the search engine responds similarly to navigational and to more severe symptom-related queries. We argue this is not the case for the following reasons: First, as demonstrated, the search engine responds with similar NDCG to all symptom-related queries. Hence, it unlikely that highly relevant results were top-ranked only for the more anxious users. Second, in navigational queries, the underlying assumption is there is only one correct answer and the user searches for it. However, in an informational query such as medical symptom search one, there is no reason to believe that the user expected to see only one result. Thus, the observed behavior in the severe-symptom queries is related to another factor. Our hypothesis is that it is related to the mental state of the user.

It may perhaps be observed without straying too far afield from our primary focus, that for the navigational queries the case was a bit different. Even though, according to the ranker, the vast majority of examples were associated with the lowest rank of 1, the computed scores were typically lower than of the baseline examples' scores associated with that rank, i.e., the scores were mostly lower than of the center of the cluster (this was the case in 73% of the examples).

4.2 Implications for Search Engine Evaluation

As previously discussed, our results suggest that the higher the level of anxiety is, the lower the probability a user would explore low ranked results. For example, as presented in Figure 2 (b), the deepest clicked result is significantly lower for queries associated with high level of anxiety. Thus, in this section we propose essential refinements to 2 common evaluation measures for search engines.

The use of $P@k$ implicitly assumes that users are interested in (or will read down to a rank of) the first k documents. NDCG implicitly assumes that users will read up to (possibly) an infinite depth of the ranked results, but that the likelihood for this decays exponentially (e.g., [31]). One of the shortcoming of $P@k$ is that we implicitly assume that the number of *displayed results* is equal to the number of *seen results*. Consequently, using this measure with a constant k value for all queries may lead to overestimation of the search engine performance. As we show below, it is important to use varying values of k, such that k corresponds to the number of seen results. A similar logic applies to the cutoff point of the NDCG calculation, where results below some k should not be taken into account. However, implementation of such policies may lead to a large overhead in terms of time and memory complexity, since it requires a separated analysis of every user-query pair. To overcome this, we propose to estimate k for every query using the trained model.

To demonstrate the importance of using a per-query value of k when evaluating the quality of search engines, we computed the $P@k$ and NDCG scores of the random sample of 65 symptom queries (as described in Section 3.5). We computed these scores for the lists of displayed results and for the partial lists of seen results. The length of the partial list was set according to the symptom mentioned in the query and the number of seen results, as computed using the screen size and the maximal y coordinate.

The average NDCG score achieved on the full list was 0.94, while the average NDCG score on the partial list was 0.89 (signtest, $P < 10^{-10}$). The average $P@k$ for the full list was 0.70, and 0.56 for the partial list (signtest, $P < 10^{-10}$). In practice, while the correlation between "full" NDCG scores and "partial" NDCG scores is good (Pearson $r = 0.9$, $P = 0.03$), it is not perfect. Therefore, for many queries, especially ones which imply a highly anxious user, we will overestimate the performance. For example, for the query "What does it mean if I have shortness of breath" the full NDCG score was 0.86, while the partial NDCG score was only 0.66. More explicitly, the SERP for that query would equate to high NDCG of 0.86, but since 90% of users asking similar queries on shortness of breath had only saw the top 6 results, the actual NDCG is 0.66.

This study shows the importance of using information on the users' level of anxiety when measuring search engine performance. These results demonstrate that a naïve measurement of relevance can cause an overestimation of the quality of search results. Thus, our results indicate that measures of relevance which considers anxiety information as well, can lead to more accurate understanding of the quality of search results, especially in cases where potentially delivering life-saving information to users.

4.3 Special Cases

We present applications of the model introduced in the previous sections in three special cases: The first is when users search for multiple symptoms. The last two are particular cases where anxiety is known to have strong effect on behavior [4, 21], namely, the

Feature	Kendall's τ
Percentage of the salient screen (height)	0.18
Rank of deepest clicked result	0.18
Number of local y minimums	0.17
Total Mouse distance	0.17
Mean y point	0.16
Max y point	0.16
Variance y point	0.14
Number of clicks below 3-rd index	0.10
Duration of session	0.09
Fraction of displayed results that were clicked	0.09

Table 3: The average Kendall's τ score while using only one feature.

anxiety level of pregnant women and the anxiety level of people asking suicide-related queries.

Multiple Symptoms. In this section we analyze search queries containing a combination of two medical symptoms. Our goal is to infer the level of anxiety of a user asking such a query. An example of such a query is when a user asks about both chest pains and headache. Intuitively, we expect that the level of anxiety of a user experiencing both symptoms to be proportional to the amount of anxiety of a user having chest pains or a headache, separately, but possibly greater than each separately.

We gathered additional queries, this time ensuring the query contains two of the medical symptoms considered. We collected 832 queries asked about 38 common combinations of symptoms (where for each combination we collected at least 10 queries). We then analyzed the predicted rank according to the user's interaction with the SERP, and compared it to the ranks of the individual symptoms. Figure 4 depicts the parameters of the linear model applied here, where the independent variables x_1 and x_2 are the min and the max ranks, respectively, and the dependent variable is the predicted rank.

Interestingly, our experiments demonstrate that the more severe symptom has a greater impact on the measured rank (i.e., the coefficient of x_1 is larger than of x_2). Namely, the level of anxiety of a user having both headache and chest pains is stronger correlated to the level of anxiety of a user having chest pains than of a user having a headache. However, the measured rank is higher than of the more severe symptom's rank, i.e., a user having both headache and chest pains is less anxious than a user having just chest pains (according to our findings), yet more anxious than one having only a headache.

As mentioned, the growth of the Internet has enabled the public to more readily access information about medical symptoms. Available websites include those that provide possible diagnoses for particular medical conditions and those that then assist people to decide whether to self-treat or consult a physician [40]. However, in this case, the healthcare assessment is based on limited knowledge of signs, symptoms, and the user's medical history. 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 [42]. An interesting application of this work, which we leave for future work, is to examine the correlation between the users' behavior, i.e., their measured emotional state, and the actual level of risk of the mentioned medical symptoms.

Variable	Slope (S.E.)	P-value
Rank of more severe symptom	0.624 (0.088)	$< 10^{-3}$
Rank of less severe symptom	0.275 (0.088)	0.003

Table 4: Model coefficients for predicting the anxiety level of queries on common combinations of medical symptoms. Model fit is $R^2 = 0.731$.

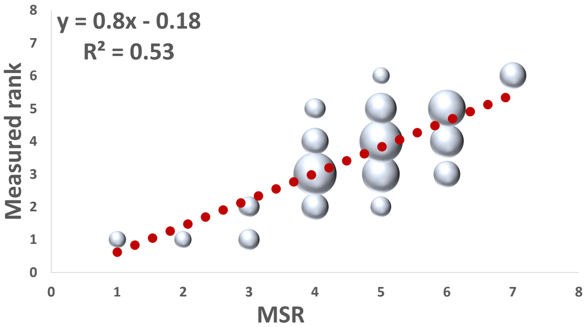


Figure 3: The effect of pregnancy on other health conditions. The size of the circles is proportional to the fraction of examples with the given MSR and measured rank.

Symptom search during pregnancy. Pregnant women have been shown to experience higher level of anxiety compared to the rest of the population [21]. Here we verify that the model is able to capture the added anxiety of pregnant women.

We collected 333 additional symptom searches, where the user asking the query also included the term "pregnant" or "pregnancy" in addition to the medical symptom, ensuring that a negation term ("not pregnant") was not used. Then, using the trained model and inference procedure, we examined the measured level of anxiety of pregnant women querying for a medical symptom.

Figure 3 shows that our findings support the hypothesis that pregnancy is correlated with a higher level of anxiety: In all cases, the average rank of queries asked by pregnant women was lower or equal to the average rank of the symptoms as rated for non-pregnant users. For example, while the rank of queries concerned with fever is 4, the average rank for such queries in pregnant women was 3. However, the model fit is moderate ($R^2 = 0.53$).

We noticed that for some of the symptoms, the measured level of anxiety was nearly equal for pregnant or non-pregnant users, while for other there was a significant difference between the measured levels of anxiety, i.e. the pregnant women were more anxious (according to the model). We therefore stratified the symptoms into two classes, according to whether or not they are typical of pregnancy as follows. We define one class of symptoms as those which are common physical symptoms during pregnancy, including nausea, fatigue, back pain, constipation, and swollen feet (according to https://en.wikipedia.org/wiki/Pregnancy), and the other class which describes symptoms not associated with pregnancy. For the first class we find that the predicted rank of the symptoms is identical to that of the non-pregnant population (the difference between ranks was ≤ 1). In the latter group of symptoms, higher level of anxiety is predicted for the pregnant population compared to the rest of the population, with a difference in ranks of 2 or more for all symptoms.

We quantify this observation using a linear regression model with two explanatory variables: the rank of the symptom, as specified

Variable	Slope (S.E.)	P-value
Rank of the symptom	0.739 (0.053)	$< 10^{-3}$
Uncommon pregnancy symptom	-1.072 (0.131)	$< 10^{-3}$

Table 5: Model coefficients for predicting level of anxiety using additional information on the prevalence of symptoms during pregnancy. Model fit is $R^2 = 0.765$.

by the medical labelers, and an indicator variable as to whether the symptom is common during pregnancy (=0) or not (=1). The parameters of this model are depicted in Table 5, showing that information on the prevalence of symptoms among pregnant women increases the fit of the model significantly (i.e., $R^2 = 0.77$), and that while pregnant women are more anxious in general (the slope is 0.739), experiencing an uncommon symptom adds to the anxiety.

Suicide-related Queries. Suicide is a preventable public health problem and a leading cause of death in the United States and many other countries [1]. The Internet offers a wealth of information for people with suicidal intentions, ranging from support groups and crisis intervention sites, discouraging individuals from committing suicide, to pro-suicide groups and how-to instructions that would not otherwise be easily accessible [36].

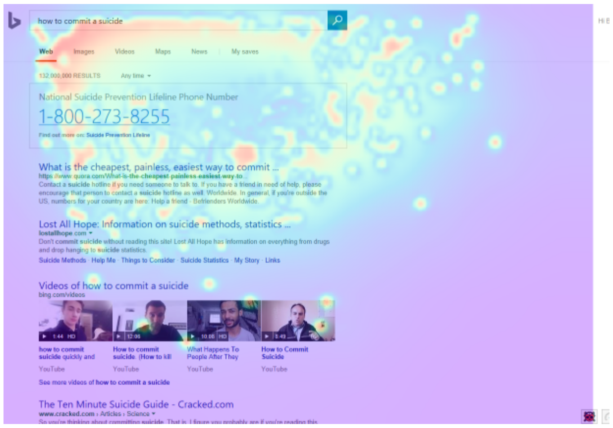
Anxiety has consistently been associated with an increase in suicidal behavior [4, 37]. In this section we aim to shed more light on the interaction with the SERP of individuals asking for practical information on how to kill themselves. Our goal is to learn about the emotional state and behavior of individuals seeking suicided related information online, and to better understand whether current interventions are useful. We collected 1375 queries where users have asked how to commit suicide. Specifically, we gathered all queries during the data period containing the text "commit suicide" or "kill myself". We then ranked these queries using the trained model, to predict the anxiety of the users asking these queries.

Our results show that in 50% of the cases the estimated level of anxiety was 2 and in 33% the estimated rank was 3 (15% had an estimated rank of 1 and all others had a rank of 4). These results suggest that users asking suicide-related queries on the Internet are highly anxious and tend to behave similarly to users asking navigational queries, meaning that they rarely explore low ranked search results. Figure 4 shows a heat map of the mouse movement during these queries. According to our data, in more than 80% of the queries, all results ranked below the index 4 (i.e., the fifth or lower ranked results) were never seen by these users. That is, any results, included websites discouraging individuals from committing suicide, ranked low by the search algorithm were completely ignored.

We note that in more than 85% of the suicide related queries we collected, a *Helpline Window* was presented to the users with the phone number of a local crisis center (i.e., in the US, the National Suicide Prevention Lifeline). Such windows (e.g., the quick answer box) are shown in only 40% of medical symptom queries. Thus, in practice, fewer than 3 results were actually been seen by most users.

These data shed light on suicidal individuals' emotional state and behavior. For search engine operators, these results highlight the critical need to provide supportive information in the highest ranked result, since lower-ranked results are rarely seen by these users.

Concluding, search engines can help save lives globally by utilizing a holistic approach to suicide prevention, including the presentation of suicide-prevention results in the upper parts of the page for particular searches. Currently, resources with harmful characteristics are frequently ranked higher than those with protective characteristics [41], and the quality of helpful suicide-related websites depends on the search terms used [8]. Efforts to improve the ranking of preventive web content seem necessary, especially when considering the portion of the screen a suicidal individual tends to explore.



National Suicide Prevention Lifeline Phone Number

1-800-273-8255

Find out more on: Suicide Prevention Lifeline

What is the cheapest, painless, easiest way to commit ...
https://www.quora.com/What-is-the-cheapest-painless-easiest-way-to...
Contact a suicide hotline if you need someone to talk to. If you have a friend in need of help, please encourage that person to contact a suicide hotline as well. Worldwide. In general, if you're outside the US, numbers for your country are here: Help a friend · Befrienders Worldwide.

Lost All Hope: Information on suicide methods, statistics ...
lostallhope.com
Don't commit suicide without reading this site! Lost All Hope has information on everything from drugs and drop hanging to suicide statistics.
Suicide Methods · Help Me · Things to Consider · Suicide Statistics · My Story · Links

Videos of how to commit a suicide
bing.com/videos

1:44 HD	12:06	10:06 HD	3:49
how to commit suicide quickly and	How to commit suicide. (How to kill	What Happens To People After They	How to Commit Suicide
YouTube	YouTube	YouTube	YouTube

See more videos of how to commit a suicide

The Ten Minute Suicide Guide - Cracked.com
www.cracked.com › Articles › Science
So you're thinking about committing suicide. That is, I figure you probably are if you're reading this.

Figure 4: Saliency map for suicide-related queries. Redder shades correspond to longer dwell times.

5 CONCLUSION AND FUTURE WORK

In this study we investigated the behavior of people during searches for medical symptoms on the Internet, and demonstrated that the severity of the symptoms, and hence the users' likely level of anxiety, is highly correlated with search behavior. We trained a ranker which can predict the implicit level of anxiety experienced by users from features extracted from the recorded interactions with the SERP. We further illustrated the importance of focusing on the top ranks of the search results page, as exceedingly anxious users tend to explore only the top of the screen.

An immediate application of this study is a refinement of search engines evaluation measures. We argue that using information on the users' level of anxiety when measuring performance is important, as ignoring it leads to significant overestimation of accuracy. Our results suggest refinements based on the learned ranker and provide a more accurate estimation of search engine performance.

Another area where our work is of importance is in the delivery of health-related information (such as information on symptoms) and especially in providing life-saving interventions. In the latter case our work shows that even simple interventions such as the help line numbers provided during suicide-related queries are seen by users. Indeed, they form the majority of the information shown to users. Thus, our work sheds light on the possibility for improving wellbeing through interventions delivered via search engines.

While the experimental results of this research are of significant value, much work is still needed to better understand the users behavior and what extensions and refinements are needed when delivering potentially life-saving information. We presume that some of the variance we could not predict may be related to demographic differences between users, people searching information for other people, and other mental or physical states of users. Future work will focus on more fine-grained analysis of these effects.

REFERENCES

[1] 2016. Suicide data. http://www.who.int/mental_health/prevention/suicide/suicideprevent/en/. (2016).

[2] 2017. Common Medical Symptoms. (2017). https://en.wikipedia.org/wiki/List_of_medical_symptoms

[3] Eugene Agichtein, Eric Brill, and Susan Dumais. 2006. Improving web search ranking by incorporating user behavior information. In *Proceedings of the 29th annual international ACM conference on Research and development in information retrieval*. ACM, 19–26.

[4] Alan Apter, R Plutchik, and HM Praag. 1993. Anxiety, impulsivity and depressed mood in relation to suicidal and violent behavior. *Acta Psychiatrica Scandinavica* 87, 1 (1993), 1–5.

[5] Ioannis Arapakis, Mounia Lalmas, B Barla Cambazoglu, Mari-Carmen Marcos, and Joemon M Jose. 2014. User engagement in online news: Under the scope of sentiment, interest, affect, and gaze. *Journal of the Association for Information Science and Technology* 65, 10 (2014), 1988–2005.

[6] Ioannis Arapakis, Mounia Lalmas, and George Valkanas. 2014. Understanding Within-Content Engagement Through Pattern Analysis of Mouse Gestures. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management (CIKM '14)*.

[7] Ioannis Arapakis and Luis A Leiva. 2016. Predicting user engagement with direct displays using mouse cursor information. In *Proceedings of the 39th International ACM conference on Research and Development in Information Retrieval*. ACM, 599–608.

[8] Florian Arendt and Sebastian Scherr. 2016. Optimizing online suicide prevention: a search engine-based tailored approach. *Health communication* (2016), 1–6.

[9] David H Barlow. 2000. Unraveling the mysteries of anxiety and its disorders from the perspective of emotion theory. *American Psychologist* 55, 11 (2000), 1247.

[10] C Botella, A García-Palacios, H Villa, RM Baños, Soledad Quero, M Alcañiz, and G Riva. 2007. Virtual reality exposure in the treatment of panic disorder and agoraphobia: A controlled study. *Clinical Psychology & Psychotherapy* 14, 3 (2007), 164–175.

[11] Israel Soares Pompeu de Sousa Brasil and Milena Pereira Pondé. 2009. Anxious and depressive symptoms and their correlation with pain severity in patients with peripheral neuropathy. *Revista de Psiquiatria do Rio Grande do Sul* 31, 1 (2009), 24–31.

[12] Andrei Broder. 2002. A taxonomy of web search. In *ACM Sigir forum*, Vol. 36. ACM, 3–10.

[13] Marc-Allen Cartright, Ryen W. White, and Eric Horvitz. 2011. Intentions and Attention in Exploratory Health Search. In *Proceedings of the 34th International ACM Conference on Research and Development in Information Retrieval (SIGIR '11)*. ACM, 65–74. https://doi.org/10.1145/2009916.2009929

[14] Mon Chu Chen, John R Anderson, and Myeong Ho Sohn. 2001. What can a mouse cursor tell us more?: correlation of eye/mouse movements on web browsing. In *CHI'01 extended abstracts on human factors in computing systems*. 281–282.

[15] Ye Chen, Yiqun Liu, Min Zhang, and Shaoping Ma. [n. d.]. Predicting User Satisfaction in SERPs with Mouse Movement Information. *IEEE Transactions on Knowledge & Data Engineering* 1 ([n. d.]), 1–1.

[16] Fernando Diaz, Qi Guo, and Ryen W. White. 2016. Search Result Prefetching Using Cursor Movement. In *Proceedings of the 39th International ACM Conference on Research and Development in Information Retrieval (SIGIR '16)*. ACM, New York, NY, USA, 609–618. https://doi.org/10.1145/2911451.2911516

[17] Fernando Diaz, Ryen White, Georg Buscher, and Dan Liebling. 2013. Robust Models of Mouse Movement on Dynamic Web Search Results Pages. In *Proceedings of the 22Nd ACM International Conference on Information & Knowledge Management (CIKM '13)*. ACM, New York, NY, USA, 1451–1460. https://doi.org/10.1145/2505515.2505717

[18] Susannah Fox. 2011. *The social life of health information, 2011.* Pew Internet & American Life Project Washington, DC.

[19] Qi Guo and Eugene Agichtein. 2008. Exploring mouse movements for inferring query intent. In *Proceedings of the 31st annual international ACM conference on Research and development in information retrieval*. ACM, 707–708.

[20] Qi Guo and Eugene Agichtein. 2012. Beyond dwell time: estimating document relevance from cursor movements and other post-click searcher behavior. In *Proceedings of the 21st international conference on World Wide Web*. ACM, 569–578.

[21] Jonathan Heron, Thomas G O'Connor, Jonathan Evans, Jean Golding, Vivette Glover, ALSPAC Study Team, et al. 2004. The course of anxiety and depression through pregnancy and the postpartum in a community sample. *Journal of affective disorders* 80, 1 (2004), 65–73.

[22] Jeff Huang, Ryen White, and Georg Buscher. 2012. User See, User Point: Gaze and Cursor Alignment in Web Search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, 1341–1350. https://doi.org/10.1145/2207676.2208591

[23] Jeff Huang, Ryen W. White, Georg Buscher, and Kuansan Wang. 2012. Improving Searcher Models Using Mouse Cursor Activity. In *Proceedings of the 35th International ACM Conference on Research and Development in Information Retrieval (SIGIR '12)*. ACM, 195–204. https://doi.org/10.1145/2348283.2348313

[24] Jeff Huang, Ryen W White, and Susan Dumais. 2011. No clicks, no problem: using cursor movements to understand and improve search. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1225–1234.

[25] Kalervo Järvelin and Jaana Kekäläinen. 2002. Cumulated gain-based evaluation of IR techniques. *ACM Transactions on Information Systems (TOIS)* 20, 4 (2002), 422–446.

[26] Jiepu Jiang, Daqing He, and James Allan. 2014. Searching, browsing, and clicking in a search session: changes in user behavior by task and over time. In *Proceedings of the 37th international ACM conference on Research & development in information retrieval*. ACM, 607–616.

[27] Thorsten Joachims. 2006. Training linear SVMs in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 217–226.

[28] Thorsten Joachims, Adith Swaminathan, and Tobias Schnabel. 2017. Unbiased learning-to-rank with biased feedback. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 781–789.

[29] Dmitry Lagun and Eugene Agichtein. 2015. Inferring Searcher Attention by Jointly Modeling User Interactions and Content Salience. In *Proceedings of the 38th International ACM Conference on Research and Development in Information Retrieval (SIGIR '15)*. ACM, New York, NY, USA, 483–492. https://doi.org/10.1145/2766462.2767745

[30] Dmitry Lagun and Mounia Lalmas. 2016. Understanding User Attention and Engagement in Online News Reading. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining (WSDM '16)*.

[31] Mounia Lalmas, Heather O'Brien, and Elad Yom-Tov. 2014. Measuring user engagement. *Synthesis Lectures on Information Concepts, Retrieval, and Services* 6, 4 (2014), 1–132.

[32] D Lekka, A Tselebis, D Bratis, G Zafeiropoulos, D Nikoviotis, A Karkanias, K Syrigos, and G Moussas. [n. d.]. 1731–The relationship between pain symptoms, anxiety and perceived family support in lung cancer patients. *European Psychiatry* 28 ([n. d.]), 1.

[33] Lori McCay-Peet, Mounia Lalmas, and Vidhya Navalpakkam. 2012. On Saliency, Affect and Focused Attention. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, 541–550. https://doi.org/10.1145/2207676.2207751

[34] Vidhya Navalpakkam, LaDawn Jentzsch, Rory Sayres, Sujith Ravi, Amr Ahmed, and Alex Smola. 2013. Measurement and modeling of eye-mouse behavior in the presence of nonlinear page layouts. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 953–964.

[35] James A O'Brien and George Marakas. 2005. *Introduction to information systems*. McGraw-Hill, Inc.

[36] Patricia R Recupero, Samara E Harms, and Jeffrey M Noble. 2008. Googling suicide: surfing for suicide information on the Internet. *The Journal of clinical psychiatry* (2008).

[37] Jitender Sareen, Tanya Houlahan, Brian J Cox, and Gordon JG Asmundson. 2005. Anxiety disorders associated with suicidal ideation and suicide attempts in the National Comorbidity Survey. *The Journal of nervous and mental disease* 193, 7 (2005), 450–454.

[38] Charles D Spielberger, Fernando Gonzalez-Reigosa, Angel Martinez-Urrutia, Luiz FS Natalicio, and Diana S Natalicio. 2017. The state-trait anxiety inventory. *Interamerican Journal of Psychology* 5, 3 & 4 (2017).

[39] Charles D Spielberger and Sumner J Sydeman. 1994. State-Trait Anxiety Inventory and State-Trait Anger Expression Inventory. (1994).

[40] Hangwi Tang and Jennifer Hwee Kwoon Ng. 2006. Googling for a diagnosisâĂŤuse of Google as a diagnostic aid: internet based study. *Bmj* 333, 7579 (2006), 1143–1145.

[41] Benedikt Till and Thomas Niederkrotenthaler. 2014. Surfing for suicide methods and help: content analysis of websites retrieved with search engines in Austria and the United States. *The Journal of clinical psychiatry* 75, 8 (2014), 886–892.

[42] Ryen White and Eric Horvitz. 2013. From web search to healthcare utilization: privacy sensitive studies from mobile data. *Journal of the American Medical Informatics Association* 20, 1 (2013), 61–68.

[43] Ryen W White and Eric Horvitz. 2009. Cyberchondria: studies of the escalation of medical concerns in web search. *ACM Transactions on Information Systems (TOIS)* 27, 4 (2009), 23.

[44] Ryen W White and Eric Horvitz. 2010. Web to World: Predicting transitions from self-diagnosis to the pursuit of local medical assistance in web search. In *AMIA Annual Symposium Proceedings*, Vol. 2010. American Medical Informatics Association, 882.

[45] Elad Yom-Tov. 2016. *Crowdsourced Health: How What You Do on the Internet Will Improve Medicine*. Mit Press.

[46] Elad Yom-Tov, Diana Borsa, Andrew C Hayward, Rachel A McKendry, and Ingemar J Cox. 2015. Automatic identification of Web-based risk markers for health events. *Journal of medical Internet research* 17, 1 (2015).