

A Study of User Queries Leading to a Health Information Website – AfterTheInjury.org

Christopher C. Yang
College of Information Science and Technology
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104
chris.yang@drexel.edu

Flaura Winston
School of Medicine, University of Pennsylvania
Center for Injury Research and Prevention
Children's Hospital of Philadelphia
Philadelphia, PA 19104
flaura@mail.med.upenn.edu

Michael A. Zarro
College of Information Science and Technology
Drexel University
3141 Chestnut Street
Philadelphia, PA 19104
michael.a.zarro@drexel.edu

Nancy Kassam-Adams
Center of Injury Research and Prevention
Children's Hospital of Pennsylvania
Philadelphia, PA 19104
nlkaphd@mail.med.upenn.edu

ABSTRACT

AfterTheInjury.org is a direct to health consumer, empirically-grounded Web-based program that is hosted by an academic medical center. Launched in 2009, it includes information and psycho-education interventional content for parents of injured children. This study analyzes search engine queries that led users to the website in order to better understand the informational needs of users of such a website. Query data were collected for the 2009 calendar year via Google Analytics, a widely used free website statistics and traffic analysis tool. Specific analyses included query length and the presence of question words. Results were compared between external and internal queries. Results from this study were put in the context of existing literature regarding user search behavior. Analyses demonstrated that external search engines were an important source of traffic to AfterTheInjury.org with 1 in 5 visits driven from the most popular search engines, Google, Yahoo and Bing. Queries were longer than typical queries (AfterTheInjury query averaged 6.27 terms; typical medical query, 2.3-2.4 terms).

Nearly 1/4 of queries leading to AfterTheInjury.org took the form of questions: 22.7% of external searches were questions with 10.7% of external queries including the term "how". The low "bounce rate" (4.2%) suggested a match between user information needs and the content on AfterTheInjury.org; however, the relatively low number of visits (7,676) given the high number of injured children per year in the US (more than 8,000,000)

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

iConference 2011, February 8-11, 2011, Seattle, WA, USA
Copyright © 2011 ACM 978-1-4503-0121-3/11/02...\$10.00

suggests that parents of injured children with informational needs may not be finding AfterTheInjury.org. These findings highlight the potential utility of query analysis to understand informational needs of users. Query behavior for users of AfterTheInjury.org differed from query behavior for those seeking general information and provide insight into how to improve dissemination strategies for similar health-related websites.

Categories and Subject Descriptors

H3.5 [Online Information Services]: Web-based services

General Terms

Measurement, Documentation, Performance, Design

Keywords

Search, Information Retrieval, Healthcare Informatics

1. INTRODUCTION

Internet users in the United States search for health and medical information at an increasing rate. The Pew Internet and American Life project's most recent estimates found that 61% of Americans looked online for health-related information in 2009 with 60% finding the online information useful [1]. Parents of injured children are an important potential audience for online health information. In 2008, more than 8,000,000 children under 18 in the US suffered an injury [2]. Injured children not only suffer physical injuries but also psychological consequences. Up to 85% of injured children and their parents develop traumatic stress symptoms (re-experiencing, avoidance, or hyperarousal) with 15–20% developing persistent and impairing posttraumatic stress symptoms (PTSS), including 5-10% who develop posttraumatic stress disorder (PTSD) [3,4]. Most develop the symptoms after medical care for the injury has ended. Parents are the most important resource for acutely injured children, and yet are often unaware or unsure of how best to help their child. After the Injury

was founded to help caregivers, particularly parents, manage medical and psychological needs of children post-injury.

1.1 AfterTheInjury.org

The Internet can provide easily accessible “just in time” information and tools to a wide range of parents of injured children for use during the months of recovery following the injury. In 2009, we launched AfterTheInjury.org to help parents help their injured child recover. Based in the empirical literature and derived from more than a decade of research, the information and psycho-education interventions on the website support the parents’ role in promoting optimal emotional recovery, with a focus on the prevention of persistent posttraumatic stress symptoms and posttraumatic stress disorder. The site includes engaging videos and interactive tools for parents. User-centered best practice design and parent feedback guided the website’s look, design, and functionality, to help ensure parent engagement. Evaluation has been incorporated throughout the process of site development and implementation [5].

In addition to continuing to enhance the website content, current efforts involve determining the most cost-effective strategies for broad dissemination of this evidence-based website as a resource to parents of injured children. To date, all of the studies related to AfterTheInjury.org have been conducted in person. This study aimed to begin to describe the behavior and informational needs of the broader audience, the web-based user of AfterTheInjury.org.



Fig 1. Screenshot of AftertheInjury.org homepage showing the process of care.

1.2 Medical Queries

The majority (66%) of online medical information seeking starts with a search engine [6]. However, this strategy results in some user dissatisfaction, as 25% report being overwhelmed by the information they found online and 22% report frustration at not being able to find the information they were seeking [6]. The need for trustworthy, understandable, and retrievable online content is an increasingly critical issue in the provision of useful medical information.

Academic medical centers are a trusted source of health information on the web [7], particularly for those from higher socioeconomic and more health-oriented groups. However, users do not always check the source of their information [6], relying rather on the highest-ranking websites in their search result. Therefore, in order to widely disseminate evidence-based health information, even credible sources such as academic medical centers need to ensure not only that they deliver excellent content but also that users can find their content when utilizing external search engines.

Making medical information retrievable has long been a major challenge. Searches for health information by laypeople differ from searches by medical professionals [8]. In a review of expert versus novice search behaviors, Bhavanani suggested general search engines were not suitable for providing authoritative information of a medical nature [9]. Non-medical language has been found to be only partially effective in finding medical information [10]. This situation creates the following challenge: how to ensure that non-experts using non-medical terminology in a general search engine find the health information they desire.

To ensure that an organization’s information is retrievable from the Web, it must be highly ranked by algorithms used by dominant search engines such as Google, Yahoo!, and Bing. The two key strategies to achieve high ranking are to ensure that the website is deemed both authoritative and relevant. To improve retrievability of information from websites, such as AfterTheInjury.org, hosted by academic medical centers (already considered authoritative), it is crucial to match user informational needs with the content on the website and ensure that the link is made by external search engines. Search engine algorithm judgments of relevancy determine the order in which results are displayed on the search engine results pages. Because it is common for general web searchers to use only the links displayed as most relevant on the first page of results, it is important that authoritative sites achieve the highest rankings possible. Recognizing that the best health information source is not always ranked most relevant by an algorithm, some search engines recently implemented programs and policies to identify and show trusted health information at the top of results screens, near where the highest ranked results normally appear. However, the choice of information sources is subjective and not all health related searches are easily identifiable, thus search engine ranking remains an important component of a website’s retrievability.

1.3 Objectives

In order to inform further development of website content and dissemination strategies, the objective of this study was to gain insight into information-seeking behavior of users of AfterTheInjury.org. As a first step, this study examines the content of queries that led users to our site for the period of one calendar year (1/1/2009-12/31/2009).

2. USER QUERY ANALYSIS

2.1 Data Extracted from Google Analytics

We captured and analyzed data in user queries collected from Jan.1, 2009 – Dec. 31, 2009 using Google Analytics, downloaded to an Excel spreadsheet for analysis. Google Analytics (<http://www.google.com/analytics/>) is the most popular website

statistics and traffic analysis tool, constituting 86.2% of the market. [11] Google Analytics was originally developed by Urchin Software Corporation, acquired by Google in 2005, and then redesigned as Google Analytics in 2006. To track data on a website, Google Analytics requires the incorporation of hidden snippets of JavaScript code on every page of the website. Each hidden snippet collects visitor data and saves it on the Google data collection servers. Processing is conducted hourly and results can be extracted at no cost on demand by website owners. The data used in this study were extracted on February 12, 2010.

2.2 Descriptive Statistics

There were a total of 7,676 visits to AfterTheInjury.org between January 1 and December 31, 2009. (See Table 1.) Most of the visitors (44.4%) came from referring websites (those that included a hyperlink pointer to AfterTheInjury.org). An additional 32.4% of visitors came directly to the site by entering the URL in their Web browser. The remaining 22.2% of visitors came from search engines such as Google, Yahoo, Bing, and others. (See Table 2) Google accounted for 80.6% of the search engine traffic to the website.

A total of 1,778 queries from these web search engines directed users to aftertheinjury.org (herein called “external queries”). These 1,778 external queries totaled 10,252 terms, resulting in an average of 5.76 terms per query. Only 75 of these queries (or 4.2% of external queries) led to no time on the site by the user (also known as the “bounce rate”). Of the 1,778 queries, 225 included the phrases, “after the injury” or “aftertheinjury.com.” As these were the name of the website, we assumed that the user mistakenly typed the URL into their search engine rather than their Web browser and the queries, “after the injury” and “aftertheinjury.com”, were omitted from further analysis. Excluding these 225 queries, the total number of external queries was 1,553 with a total of 9,731 terms for a corrected average of 6.27 terms per query.

On-site searches totaled 404 queries with 1,061 terms, for a mean of 2.6 terms per query. There were 207 blank queries that were set to the default text in the search form on the website, “search this site”. Removing these blank queries gave a corrected total of 197 queries with 443 terms for a corrected average of 2.2 terms per query.

Table 1. Number and percentage of visits from direct traffic, referring sites, and search engines to AfterTheInjury.org

	Number of Visits	Percentage
Direct Traffic	2,491	32.4%
Referring Sites	3,416	44.4%
Search Engines	1,713	22.2%
Total	7,676	100.0%

Table 2. Number and percentage of visits from individual search engines to AfterTheInjury.org

	Number of Visits	Percentage
Google	1,327	17.3%
Yahoo	223	2.9%
Bing	52	0.7%
AOL	36	0.5%
MSN	24	0.3%

Table 3. Number of queries from external searches (i.e. search engines) and on-site searches

	Number of Queries	Term Count	Terms per Query
External Searches	1,778	10,252	5.76
Corrected external Searches ^a	1,553	9,731	6.27
On-Site Searches	404	1,061	2.6
Corrected On-Site ^b	197	443	2.2

2.3 Queries Made as Questions and Their Classifications

The majority of external queries involved strings of words or phrases. However, 353 queries (or 22.7% of external queries) involved specific questions (57 queries contained question words and question marks and the remainder only contained question words). A total of 3,344 terms in the 353 question-type queries gives an average of 9.5 terms per query. By comparison, non-question-type queries included a total of 6,387 terms for an average of 5.3 terms per query. Only six on-site queries were in question form, accounting for 1% of all internal queries. On-site question-type queries had an average of 7.16 terms per query.

Among the question-type queries, “how” was the most common starting question word, accounting for almost one half (167) of all question-type queries (47.3% of question-type queries, 10.7% of all external queries). An additional 63 question-type queries began with “what” (17.8% of question-type queries, 4.1% of all external queries): “what happen/happens”, “what is”, “what if”, “what should”, “what to”, or “what type”. Of the “how” question-type queries, 65 (38.9%) started with “how to” such as “how to tell if a stress reaction is getting worse”, “how to make a child sleep alone after traumatic experience”. An additional 66 (39.5%)

^a Removed 225 instances of URL name “after the injury” or “aftertheinjury.org”

^b Removed 207 instances of default search value “search this site”

searching behavior will be an important strategy to drive relevant traffic to AfterTheInjury.org.

3.3 Finding Health Information

Queries may lead users to the website who then did not find relevant information in the website. The low bounce rate suggests that users who arrived at AfterTheInjury.org found information of interest. However, the relatively low number of users given the large number of injured children in the US each year suggests that many parents of injured children who may be seeking information do not find this site.

Finding trusted information from an authoritative source continues to be an issue in online healthcare searching. Several organizations and search engines have come to recognize the need to assist healthcare searchers find trustworthy sources of information. One such group is Health on the Net (HON), a non-profit organization that “promotes and guides the deployment of useful and reliable online health information” [13]. Sites meeting criteria, including After the Injury, display the HON icon on their pages indicating they are a source of trusted information.

Google recently implemented a policy of showing information from trusted sources near the top the search results screen. The policy was described as offering one source from a governmental health source, a medical institution, and a commercial source [14]. At present this includes: Mayo Clinic, Medline Plus, and WebMD, in addition to Google Health. The search engine Bing, owned by Microsoft, also promotes information from sources they judge to be reliable, including Mayo Clinic, the American Cancer Society, MD Consult, and Gold Standard [15]. These efforts demonstrate the desire in industry to provide high quality healthcare search results.

The methods used in this study provide a cost-effective framework for future testing of strategies to improve the relevance of content on and dissemination strategies (including search engine optimization) for AfterTheInjury.org. These strategies should take into account the limitations of Web searching, both in terms of the current algorithms used by search engines and user advice-seeking behaviors. Search engine algorithms rely heavily on keyword matching to retrieve relevant web pages. The vocabularies used by health consumers often differ from those used by health professionals who create Web content [12], which reduces the match between the information sought and websites recommended by search engines, particularly for health related information. This underlines the need for website authors to utilize clear lay language not only to optimize readability and user comprehension, but also to help potential users find their site. Current developments in search engine algorithms extend strategies based on keyword matching to also include link analysis and query log analysis to improve the ranking mechanisms.

Online healthcare information is spread across many web pages and websites. In this study, many queries led visitors to a glossary page that contains a comprehensive listing of terms relevant to the information on the site. Because the glossary page has many potential keywords listed, keyword matching may direct users to this page over more relevant pages on the website. An information seeker is likely to require visits to several sources in order to gain a comprehensive understanding of an information need [16]. This distribution of information across web pages and

websites means that providers of healthcare information could serve their audience by collaborating with related providers of resources.

Healthcare information seekers might not find all the information they need if they visit only websites that answer a specific question. Rather, searchers need to form a strategy to find comprehensive information and view sources of both general and specific information. A limitation of query log analysis is the lack of context of the user’s search. There is no indication of their search strategy, nor do we know if they have visited other, possibly complementary, websites.

After the Injury is designed in such a way as to guide the visitor through several stages of the post-injury process. The site, designed in a user-centered fashion is meant to provide easy to use information. A core component is the creation of a step-by-step navigation process. It is intended that visitors utilize resources in a prescribed order, to support their learning objectives. The low number of on-site searches suggests that users are having some success utilizing the navigation to find the information they desire. This is encouraging, as often information must be viewed in context to provide the best outcome for the patient.

3.4 Limitations

The results of this study were limited by our methods which analyzed only queries that actually led users to AfterTheInjury.org. The query behavior reported may not be generalizable to all users seeking information about child injury. Further, arriving at the website is not only a function of the user’s query but also of the algorithms used by search engines and so may not be completely representative of users’ information-seeking behavior. Future studies should examine query logs from search engines to identify missed opportunities to link user information seeking with available content on AfterTheInjury.org.

3.5 Future Work

In future work we intend to study techniques to analyze query types and provide ways to direct searchers to the page that best satisfies their need. Because there is a large network of medical information sites that serve the public, the potential exists to provide cross-linking services with partner websites. For example, if a search engine leads a searcher to the After the Injury website for adult trauma needs, After the Injury may be able to detect that specific need based on the terms in the query and provide a link to a site that might better serve the information need of the searcher.

4. IMPLICATIONS AND CONCLUSION

Search engines are an important source of traffic to health information websites. Analysis of search queries can provide an indication of the information sought by users of the websites. Given the current limits of both user searching behavior and search engine algorithms, it is incumbent upon website authors to review user information needs in order to optimize their content so that a broader audience can be reached.

Our analysis indicates queries that led to the After the Injury website, with an average length of 6.27 terms per query, were considerably longer than those reported in previous studies. Queries in question form were even longer, averaging 9.5 terms

per query. This suggests the site provides information for visitors searching for particular information. The site attracts a larger percentage of “how” questions than any other, particularly “how long” and “how to” types of questions.

A key motivation of our study is to understand how users reach the site in order to identify areas where we may enhance the content and site structure for two reasons. First, to serve the information needs of parents caring for injured children, and second, to increase the relevancy rankings in general web search engines. These efforts are related, as information is not useful to a person if it is not found. Given current information seeking practices of healthcare consumers, we hope this study is a step towards finding new ways to get authoritative, trusted, and useful information to healthcare information seekers.

5. REFERENCES

- [1] Fox, S. 2009. The Social Life of Health Information. Pew Internet & American Life Project (June 2009). <http://www.pewinternet.org/Reports/2009/8-The-Social-Life-of-Health-Information.aspx>
- [2] Injury prevention & control: data & statistics (WISQARS™), <http://www.cdc.gov/injury/wisqars/nonfatal.html>, retrieved 3/9/2010.
- [3] Kassam-Adams, N. and F. Winston. 2004. Predicting child PTSD: The relationship between ASD and PTSD in injured children. *Journal of the American Academy of Child and Adolescent Psychiatry*. 43, 4 (April 2004), 403-411.
- [4] Kassam-Adams, N., C. Fleischer, and F.K. Winston. 2009. Acute stress disorder and posttraumatic stress disorder in parents of injured children. *Journal of Traumatic Stress*. 22, 4 (Aug. 2009), 294-309.
- [5] Kassam-Adams, N., Marsac, M., & Winston, F. K., 2009. AfterTheInjury.org: Development of a website for parents promoting secondary prevention of child traumatic stress. *Proc. NATO Advanced Research Workshop in Human Societal Dynamics (Internet and Mental Health)* (Montreal. 2009).
- [6] Fox S. 2006. Online Health Search 2006. Pew Internet & American Life Project, October 2006. <http://www.pewinternet.org/Reports/2006/Online-Health-Search-2006.aspx>
- [7] Dutta-Bergman M. 2003. Trusted Online Sources of Health Information: Differences in Demographics, Health Beliefs, and Health-Information Orientation. *Journal of Medical Internet Research*. 5, 3 (Jul. – Sep. 2003). <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550562/>
- [8] Spink A, Yang, Y, Jansen, J, Nykanen, P, Lorence, DP, Ozmutlu, S, Ozmutlu, H C. 2004. A Study of Medical and Health Queries to Web Search Engines. *Health Information and Libraries Journal*. 21, 1 (Mar. 2004), 44-51.
- [9] Bhavnani S K. Domain-specific search strategies for the effective retrieval of healthcare and shopping information. *Proc. CHI '02: CHI '02 Extended Abstracts on Human Factors in Computing System*. (2002), 610-611.
- [10] Lau T, Horvitz E. Patterns of search: analyzing and modeling web query refinement. 1999. *Proc. the Seventh International Conference on User modeling* (1999), 119-128.
- [11] Usage of traffic analysis tools for websites. W3Techs, http://w3techs.com/technologies/overview/traffic_analysis/all, retrieved 3/9/2010.
- [12] Zeng, Q., Goryachev, S. Tse, T. Keselman, A., and Boxwala, A. 2008. Estimating consumer familiarity with health terminology: a context-based approach. *J. AM Med Inform Assoc*. 15, 3 (May-June 2008), 349-356.
- [13] Health On the Net Foundation, <http://www.hon.ch>, retrieved 8/20/2010.
- [14] Fox S. 2010. Health Sites: Some Are More Equal Than Others, <http://www.pewinternet.org/Commentary/2010/January/Health-Sites-Some-Are-More-Equal-Than-Others.aspx>, retrieved 8/20/2010.
- [15] Bing Community Blog: User Needs, Features and the Science behind Bing, <http://www.bing.com/community/blogs/search/archive/2009/06/01/user-needs-features-and-the-science-behind-bing.aspx>, retrieved 8/20/2010.
- [16] Bhavnani S K., Peck, F A. Scatter matters: Regularities and implications for the scatter of healthcare information on the web. *Journal of the American Society for Information Science and Technology*. 61, 4 (2010), 659-676.