

# PETTT TECHNICAL REPORT

## PETTT-01-AS-03



## Improving Search Engine Position of Internet Educational Materials: Design Heuristics and Indexing Methods

Aaron J. Louie, Jacob S. Burghardt, Ralph Warren, Jr.,  
Scott K. Macklin, Fredrick A. Matsen  
February, 2000

### Abstract:

The Internet provides a readily accessible educational resource for individuals outside the context of defined curricula. These “learners-at-large” often use search engines to locate educational materials that meet their needs. It is necessary for creators of educational content on the Web to understand the factors affecting the ability of search engines—and, consequently, learners—to locate their site.

In our study, we determined that (1) search engines use different strategies for ranking sites, (2) search engine positioning can be optimized by following heuristics for site organization, content design, and submission strategy, and (3) search engine position is related to the rate at which a site is accessed. We suggest that these observations may be relevant to those creating content on the Web for learners-at-large.

To cite this document, use the format:

Turns, J. and Wagner, T. (2001). Continuing Medical Education: Observations of CME Course on Feb. 8 and 9, 2001. Technical Report PETTT-01-PT-01. Program for Educational Transformation Through Technology, University of Washington, Seattle, Washington.

*Notice: PETTT has made its working papers, reports, and other documents available for your free use, but only if you (a) reference us in your bibliography and (b) tell us about it by emailing us at [petts@u.washington.edu](mailto:petts@u.washington.edu).*

## Introduction

### **Helping “Learners-At-Large” Locate Educational Resources On The World Wide Web**

The Internet provides a readily accessible educational resource for individuals outside the context of defined curricula. These “learners-at-large” often use search engines to locate educational materials that meet their needs. It is necessary for creators of educational content on the Web to understand the factors affecting the ability of search engines—and, consequently, learners—to locate their site. We present an empirical study of the design modifications of the Arthritis Source, a health care Web site providing information for learners-at-large, and report the effect of these modifications on search engine positioning. We describe strategies and techniques that educators may use to enable potential users to locate their Web sites. Search engine positioning refers to gaining a superior relative search engine position of a Web site compared to other Web sites.

Instructors often use the Web as a method for communicating syllabus materials to students enrolled in specific courses. By contrast, this article concerns the provision of educational content to a different type of learner, the learner-at-large, who turns to the Web to seek educational materials outside of an assigned curriculum. These learners can be classroom students wishing to go beyond the standard course content, professionals conducting self-guided continuing education, or patients at home wishing to learn more about surgical treatment options for severe rheumatoid arthritis.

Such learners use the Web to search for information that satisfies their learning goals. Norman and Spohrer (1996) discussed “learner-centered education” in terms of self-motivated learners seeking knowledge and skills in order to solve particular real-world problems. These authors proposed that learners are often searching for answers to specific questions as those questions arise. In this way, the learning accomplished by learners-at-large often has a situated element that is congruent with the task-based learning approach advocated by educational researchers such as Seedhouse (1999), Whittington (1998), and Starr (1997).

With the expansion of the Web, educators have the potential to reach unprecedented numbers of learners. Shneiderman (1998) proposes that the donation of educational Web sites to the public should be a major focus of teaching and learning on the Web. This has quickly become the case for health care information on the Web in the past several years. Kinzie et al. (1996) created “Netfrog,” a Web site that allowed learners to dissect a virtual frog, and tracked how the site was accessed by studying the Web server’s log files. Log file analysis revealed that only 26.6% of the unique domain names accessing the site during that time could be identified as belonging to U.S. educational institutions. This statistic suggests that many of the learners that accessed “Net Frog” may have been learners-at-large.

A key challenge for learners-at-large who are turning to the Web is the task of finding resources that meet their learning needs. Some of the methods that a learner-at-large can use to locate an educational Web site include (a) searches of known educational Web sites, (b) personal references, (c) references from other sites or organizations, (d) advertisements, (e) search engines, and (f) educational resource “gateway” sites.

As more instructional resources become available on the Web, educators in K-12, post-secondary and professional programs can benefit from organized directories of quality resources on the Internet. Federal agencies and academic institutions have begun to support gateway Web sites that organize access to peer-reviewed Web-based educational resources. The GEM project and the Merlot project are examples of gateway Web sites that aim to satisfy this need for educators. The U.S. Department of Education’s Gateway to Educational Materials project (<http://www.thegateway.org/>) provides access to a wide range of un-catalogued educational materials available on federal, state, university, non-profit, and commercial Web sites. The peer-reviewed Multimedia Educational Resource for Learning and Online Teaching (<http://merlot.cdl.edu/Home.po>) is another gateway-type resource designed by and developed for faculty in higher education.

Since the late 1990s, government agencies and professional societies have recognized that easy access to accurate and reliable health information on the Internet was lacking. Several initiatives and projects have been funded to build and maintain portal or gateway Web sites directing the public to quality health

information. Examples of these initiatives in the health care area include Healthfinder, MEDLINEplus, and the Medical Matrix Project.

Developed by the U.S. Department of Health and Human Services in 1997, Healthfinder (<http://www.healthfinder.gov/>) is a clearinghouse for government, academic and non-profit Web sites in the basic and applied health sciences, enabling access to online publications and databases. MEDLINEplus (<http://www.nlm.nih.gov/medlineplus/aboutmedlineplus.html>) is a recent Web-based project of the National Library of Medicine at the National Institutes of Health. It includes extensive information about specific diseases and conditions (including clinical trials) with links to medical dictionaries, lists of hospitals and physicians, and to health information in Spanish and other languages. The Medical Matrix Project (<http://www.medmatrix.org/>) is sponsored and managed by an inter-professional association of health care professionals, the American Medical Informatics Association's Internet Working Group. An editorial board ranks health education resources available through the Internet based on overall quality of the content, multimedia features, and usefulness to clinicians. The Medical Matrix Project was developed for United States physicians and health care workers but it is freely available and accessible to the general public.

### ***Finding Content On The Internet Through Search Engines***

The gateway projects are becoming more widely known among educators, but learners-at-large are less likely know of them. It is likely that, when a patient seeks information concerning an illness or health care options, the search engine serves as a primary resource. Unfortunately, educational gateway sites are unlikely to be among the first set of Web sites returned by a search for educational content on the most commonly used search engines.

Search engines can steer a learner-at-large to the answers to his or her questions with a simple query. In the context of locating educational Web sites, a learner must know enough about their topic to enter relevant keywords, and must filter through the search results to find the desired content, if it is available. Once a site has been found, a learner can use a “bookmark” or their Web browser's history mechanism to revisit it (Tauscher & Greenberg, 1997). The problem with search engines from the learner's perspective is that the most relevant URLs (Web site addresses) for their learning goals may appear at the end of a long list of search results. This dilemma is especially apparent when educational materials on a given topic are in direct competition with commercial sites. For the learner-at-large, this means that the top ranked sites (often several pages of search results) may not contain the learning materials that they are looking for. For an educator producing content who wishes to have her content found by learners-at-large, search engine position becomes a matter of substantial importance.

Learners-at-large vary widely in their search skills as they approach the task of locating useful information with search engines. Hill (1999) identified three types of users of open-ended information systems: (1) naive, (2) somewhat knowledgeable, and (3) knowledgeable. Naive users may often have difficulty adapting previous search behavior to successfully inform new search decisions. Knowledgeable users, however, are able to integrate new feedback at each phase of the search process, a critical feature for successful use of search engines on the Web.

In addition, it is not clear whether current search engines provide useful results and assistance for learners-at-large, who possess a wide range of goals and motivations. In an exploratory study of hypermedia navigation, Barab, Bowdish and Lawless (1997) identified distinct profiles of navigation behavior. Some users may be motivated primarily by learning goals while others may be motivated more by performance goals. Their results suggest that multimedia and other features in the environment may distract some users, while others may quickly give up exploration in an unstructured hypermedia environment.

Search engine sites do not typically provide explicit support for identifiable goals, motivations and prior search skills for their users. As a result many learners-at-large may haphazardly retrieve and explore Web-based educational content when using search engines. This compounds concerns about the quality of content that naive searchers may locate when they seek health information on the Internet.

It is apparent that the use of the Internet is increasing among patients seeking to learn more about their condition (Hardey, 1999, and Dyer, 1998). McCullough (2000) estimated in April 2000 that seventy-five million Americans had access to the Internet and more than half of these individuals sought health information online at least once per month. The Internet has thus become a significant part of how patients come to learn about their health concerns and needs, often impacting the patient-doctor relationship.

Bader and Braude (1998) noted, “Patients anxious to participate in decisions about their own treatment have turned to the Internet to confirm diagnoses, validate physician-recommended treatment, or

seek alternative therapies.” But this participation may not always lead to advantageous results: both the company selling treatments and the educator who wishes to inform patients about how they can manage their conditions are often competing for the attention of the same population of learners. Soot, Moneta, and Edwards (1999) used five common Internet search engines to locate information on four varieties of vascular surgery. They found that 65.8% of the sites “had no useful patient-oriented information.” Looking only at the 33.2% of sites that were categorized as being relevant for patients, “one third of the information” was deemed “misleading or unconventional.”

Beredjikian, Bozentka, Steinberg, and Bernstein (2000) evaluated the quality of orthopedic content on the Internet and raised significant concerns about: (a) the likelihood of retrieval of health related Web sites by search engines, and (b) the quality of medical information found. They searched for the phrase “carpal tunnel syndrome” on the five most commonly used search engines and found that of the 250 Web sites (the first fifty sites identified by each search engine), 175 had a unique URL and seventy-five were duplications. Surprisingly, not one Web site was identified by all five search engines and only two sites were listed by four of the five search engines. They reported that, for Web sites found by the search term “carpal tunnel syndrome,” less than half of provided “conventional” medical information and twenty-three percent offered unconventional or misleading information. This raises questions about the adequacy of coverage of the search engines for health information, and reinforces a frequently cited finding: any one search engine has limited coverage of the entire Web, probably no greater than one-third of the “indexible Web” (Lawrence and Giles, 1998).

Other studies report positive findings about the quality of health information on the Web and the impact on learners-at-large. Leaffer and Gonda (2000) studied senior citizens who were taught how to conduct health information searches on the Internet. The resulting pattern of Internet use and related effects on the treatment relationship were noteworthy: “Two thirds of those who searched for health information on the Internet talked about it with their physicians, with more than half reporting they were more satisfied with their treatment as a result of their searches and subsequent discussion with their physicians.” There is some evidence that suggests that not only are patients more satisfied with their treatment when they have access to information about their conditions, but that the education itself is therapeutic. In a meta-analysis of 76 studies on the effects of arthritis education, Lorig, et al. (1987) found that 61% of patients had clinical improvements as a result of health education. These results underscore the importance of high-quality, easily accessible educational materials that are designed to optimize their rankings in major search engines.

## ***Research In Search Engine Positioning***

Tunender and Ervin (1998) investigated the effects of promoting a Web site created at the University of Missouri-Columbia. The authors placed experimental character strings in the title tag, meta description tag, and throughout the text of all pages on their site. The site was then submitted to 5 frequently used search engines. They found that, after varying amounts of time, the pages could be found in four out of the five engines by querying for the experimental character strings. While not all of these experimental keywords were accessible in any of the monitored engines, five of the eight character strings were accessible using Excite (<http://www.excite.com>) by the 23rd day of the study. Although this method of search engine positioning was not entirely successful, it demonstrates that educational Web sites can be promoted in search engines that may be commonly used by learners-at-large.

Based on the literature discussed herein and on the experience of the authors, we maintain that search engines have methods of discovering and ranking Web sites that are consistent and predictable. This claim, in conjunction with the findings of Tunender and Ervin (1998), leads to the hypothesis that the search engine position of educational Web sites can be optimized with informed design and periodic submissions. While it may be true that not every search engine assigns rankings of Web sites in the same way, we hypothesize (1) that, by following contemporary design heuristics, educational Web-designers can improve the search engine position of their sites in several of the most frequently used search engines. A second hypothesis is (2) that improvement in the ranking of a keyword in search engines’ indexes will be correlated with increases in hits to the page associated with that keyword. This substantiates the notion that search engines are a key means of finding Web sites. As there is often persistent competition for ranking within searches for keywords, a third hypothesis is (3) that search engine position of a particular keyword will degrade over time, warranting periodic resubmission.

Before testing these hypotheses with design interventions on three patient education Web pages, we will first provide background on the mechanics of search engines and contemporary design theory for optimizing the search engine position of Web sites. To help determine strategies for optimizing the visibility of academic content on the Web, the University of Washington's Program for Educational

Transformation Through Technology<sup>1</sup> has implemented three patient education Web pages as test beds for researching the relationship between a site's design and its search engine position (<http://depts.washington.edu/pett>). Here, we present our use of these three pages to test the above hypotheses, which are central to the positioning of Web education content for learners-at-large.

## Design Heuristics

### ***Basic Search Engine Varieties***

To provide a working understanding of search engine mechanics, we will define a taxonomy of search engines, with special attention to the variety of engine that engaged by our proposed design methodology: the “robot” or “crawler.” Search engines can be separated into three basic varieties, based on their method of finding and ranking sites: “directory” engines, “robot” or “spider” (“bot”-based) engines, and “hybrid” engines (adapted from [www.searchenginewatch.com](http://www.searchenginewatch.com)). Directory engines, such as OpenDirectory (<http://www.opendirectory.com>), depend on people to assemble the rankings. A Web page's designer submits a short description of their entire site to a directory engine. The engine searches for matches to a user's query based on the short descriptions that have been accepted and categorized by human reviewers, not on the sum of the content on a page or in a site. Bot-based engines, on the other hand, create listings automatically, without the individual attention of a person. Computer programs called “robots”, “spiders”, or “bots” continuously roam the Web, using procedural algorithms to collect information found on Web sites. With this variety of engine, the search results that users see are based on the information amassed by the bot's algorithms. Hybrid search engines, such as MetaCrawler (<http://www.metacrawler.com>), include a mixture of directory and bot characteristics, combining search results from other engines and maintaining an associated directory.

Though not all bot-based search engines operate in the same way, they do share three basic elements of functionality (<http://www.washington.edu/catalyst>). First is the actual robot, the computer program that is designed to roam the Web visiting Web sites and following the links within those sites. These agents scan through a Web site, following their predetermined algorithms to look for relevant information in the site's content, which is then stored in another of the common elements of bot-based engines: the search index. Many bots use a text-based crawling method, so they cannot “see” many of the more advanced features on the Web (such as frames, JavaScript, or Flash animations). A summary of Web page metadata visible to common bots is presented in Table 1, below. The second shared component is the search engine's index. An engine's search index is a database of all the data that the bot has collected from each Web page it has catalogued.

(Insert Table 1 here)

### ***Design Heuristics For Improving Search Engine Position***

Great effort may be expended putting educational materials online, but, if few learners find those materials, an educator's work can be in vain. Even if educational Web designers submit their sites to search engines in an attempt to bring more learners to their pages, they may not design those pages with the bot's measures of merit in mind. Without a working knowledge of current design heuristics, educators may find their sites under-ranked and under-visited. Educational Web designers need not accommodate the specific needs of all search engines—over 90% of the search engine traffic to most Web sites originates from 8 to 10 major search engines (<http://www.webposition.com>).

Based on the above references, we propose a list of design heuristics for educators who wish to improve the search engine position of their pages. These design heuristics are an amalgamation of the information presented in the above-cited Web sites, and, as such, individual references are not included with each design suggestion. It is important to note that we are not proposing a definitive list of design suggestions. The design heuristics presented here represent a compilation of available resources on the subject and do not necessarily include all the features that are attractive to bots (<http://www.webdevelopersjournal.com>, [www.webposition.com](http://www.webposition.com), [websearch.about.com](http://websearch.about.com), [www.searchenginewatch.com](http://www.searchenginewatch.com), [www.builder.com](http://www.builder.com), [depts.washington.edu/catalyst](http://depts.washington.edu/catalyst), [searchengineposition.com](http://searchengineposition.com)).

---

<sup>1</sup> The Program for Educational Transformation Through Technology (PETTT) seeks to enhance the effectiveness of the University of Washington by creating a campus framework to promote the thoughtful exploration, development, assessment, and dissemination of next-generation technologies and strategies for teaching and learning.

We outline three general categories of design heuristics that educators can follow to create Web sites that will be attractive to bot-based search engines. These include (1) meta tag heuristics, (2) page design heuristics, and (3) consistency heuristics. Table 1 contains code that illustrates these heuristics.

### ***Meta Tag Heuristics***

Educational Web designers must carefully consider the keyword and description meta tags included in their pages, as these tags are the primary data that search engines use to categorize a Web page.

- **Keywords:** Ideally, each page should have about eight meta tag keywords, around 20 characters each, that describe the content of the page. Robots look at the number of times a particular keyword appears in the list of meta tags (the meta keywords field, see Table 1) and in the visible, on-screen text (<http://websearch.about.com>).
- **Description:** Each page should also contain a descriptive tag of 20 words or less. This tag is usually a single sentence, and it may be beneficial to echo the topic sentence of the page's first paragraph (<http://www.searchenginewatch.com>).

### ***Page Design Heuristics***

Page design heuristics refer to the length of various parts of a page's content. Search engines thoroughly inspect the content of a Web page to determine how it should be ranked for a variety of keywords. There are several rules that should be considered in the design of pages to improve their search engine position:

- **Page length:** Web sites should be divided into pages that are about 250 words long. This is very much dependent on the search engine for which the site is designed. Altavista suggests a range of 5-10kb of text on each page, which translates to approximately 850 words. Other search engines, such as Google, do not have such strict requirements. Some search engine "crawlers" will reduce the ranking of sites with longer pages (<http://www.webposition.com>).
- **Title length:** Each page should have a title that is about 40 characters long, and should be used as heading text on the actual page as well (<http://www.builder.com/Business/Promote/>).
- **Topic sentence length:** Each page should ideally have a topic sentence that is about 20 words long, where a topic sentence is defined as the first non-heading text on the page. Some search engines use this information as a description of the page, much like a meta tag.

### ***Consistency Heuristics***

Consistency heuristics are concerned with the consistency of meta tags and textual content throughout the site. Some search engines give sites better rankings if they use consistent terms throughout a Web site. Although not all engines may use this measure, it is an important factor to take into consideration if an educator wishes to ensure the best possible rankings on a number of search engines. Consistency should be observed in three areas:

- Consistency between the title and meta tag of pages with related content unifies their content into a network of related terms.
- Consistency between the meta tag words and the content of pages provides unity between the categorization of pages and their material.
- Consistency between the meta tag and textual content of different pages on the site increases the relevancy score given to the entire site by the bot (<http://www.webposition.com>).

### ***The Submission Process***

Once a Web site has been designed and meta tags have been applied, it should be submitted to search engines. Tunender & Ervin (1998) report empirical evidence of bots visiting submitted educational materials within 17 days of their submission. The submission process typically involves manually accessing each search engine's Web site and submitting the site for consideration though services and software applications have been developed to automate the process.<sup>2</sup>

---

<sup>2</sup> Submission services: [www.webposition.com](http://www.webposition.com), [www.searchenginewatch.com](http://www.searchenginewatch.com), [www.positionagent.com](http://www.positionagent.com), [www.submitwolf.net](http://www.submitwolf.net), [www.submitcorner.com](http://www.submitcorner.com)

## Methods

### ***Experimental Test Of Design And Submission Strategies***

To assess the effectiveness of the design heuristics and submission processes described above, a subset of the design heuristics were applied to three patient information Web pages on orthopedic surgery created at the University of Washington Medical Center. These Web pages had been online prior to this experiment, and thus the experimental manipulations described herein were “design interventions” to improve the ease of access for potential learners-at-large. Two separate design interventions were made, with the second intervention being a continuation of the first. We predicted that, by altering the design of the experimental pages in accordance with our heuristics and by submitting the redesigned pages to search engines,

1. search engine position of individual keywords of the experimental pages would improve over time and that this effect would be more pronounced when a greater number of design the heuristics are implemented,
2. changes in the ranking of the most representative keyword for each of the experimental pages would be correlated with increases in hits to those pages, and
3. improvements in search engine position of a particular keyword would be lost over time as other sites are assessed and ranked by the search engine.

### ***Materials***

The experimental design interventions were conducted on three pages (Dislocat.htm, RotCufSu.htm, and ShoulRep.htm) of the “Shoulder Source” (<http://www.orthop.washington.edu/shoulder/shoulder.htm>)<sup>3</sup>, which itself is a subset of University of Washington Arthritis Source Web site (<http://www.orthop.washington.edu>). The Arthritis Source is a Web-based information resource that provides patient education materials about arthritis to learners-at-large. It is sponsored and maintained by the University of Washington Department of Orthopedics and Sports Medicine, and provides learners-at-large with access to information about the many forms of arthritis. It also provides strategies for living with this life-long disease to patients, their family members, students, health care professionals and the general public.

Analysis of the log files for the Arthritis Source revealed that the site had been accessed from a variety of domain names, including people located at the University of Washington hospital, other medical schools, and home users. The latter group has the greatest potential to be the learners-at-large described in the introduction. For the purposes of this experiment, patient information pages describing the procedures and risks of three different orthopedic surgeries served as a test-bed for measuring the effectiveness of interventions intended to optimize search engine position

In their original design ([http://www.orthop.washington.edu/shoulder/Dislocat\\_0.htm](http://www.orthop.washington.edu/shoulder/Dislocat_0.htm), [RotCufSu\\_0.htm](http://www.orthop.washington.edu/shoulder/RotCufSu_0.htm), [ShoulRep\\_0.htm](http://www.orthop.washington.edu/shoulder/ShoulRep_0.htm)), each of these three Web pages consisted of a single page, several screens in length—a principle that violates the page design heuristics stated above. Each page had been initially designed with meta tag keywords, but the selected keywords did not conform to the design heuristics described above.

(Insert Figure 1 here)

### ***Data Collection***

To assess the effectiveness of the two design interventions on the search engine position of each of the three experimental pages, two metrics were used: hit frequency and keyword search engine position.

---

<sup>3</sup> Shoulder Dislocation Repair Surgery: [http://www.orthop.washington.edu/shoulder/Dislocat\\_0.htm](http://www.orthop.washington.edu/shoulder/Dislocat_0.htm), [Dislocat\\_1.htm](http://www.orthop.washington.edu/shoulder/Dislocat_1.htm), [Dislocat\\_2.htm](http://www.orthop.washington.edu/shoulder/Dislocat_2.htm); Rotator Cuff Repair Surgery: [http://www.orthop.washington.edu/shoulder/RotCufSu\\_0.htm](http://www.orthop.washington.edu/shoulder/RotCufSu_0.htm), [RotCufSu\\_1.htm](http://www.orthop.washington.edu/shoulder/RotCufSu_1.htm), [RotCufSu\\_2.htm](http://www.orthop.washington.edu/shoulder/RotCufSu_2.htm); Total Shoulder Replacement Surgery: [http://www.orthop.washington.edu/shoulder/ShoulRep\\_0.htm](http://www.orthop.washington.edu/shoulder/ShoulRep_0.htm), [ShoulRep\\_1.htm](http://www.orthop.washington.edu/shoulder/ShoulRep_1.htm), [ShoulRep\\_2.htm](http://www.orthop.washington.edu/shoulder/ShoulRep_2.htm)

## Hits Per Day

This metric is the number of times that a particular URL is requested in one day. This value was tracked by consulting the [www.orthop.washington.edu](http://www.orthop.washington.edu) Web server's log files, a record of all of the data transactions that the Web server has made. Hit frequency data was collected daily and analyzed with FastStats (<http://www.mach5.com/fast/>), a commercial software application. The baseline value for daily hit frequency to each of the three pages before the design interventions was calculated as the average hits per day from an extensive set of data that began in January 1999 and ended in December 1999. FastStats counted multiple requests for the same URL from the same IP address as multiple hits from a single user. Thus, the number of discrete IP addresses that accessed the site was measured as the number of discrete users that visited that site.

## Search Engine Position

The ranking of the experimental pages varied depending on the search engine being used and the keyword being used as the query. Tracking this variation involved monitoring many components: each page had nine keywords and search engine position for each keyword was recorded in 15 different engines, resulting in 405 search engine positions being monitored during the course of the experiment. Keyword ranking data for each page was collected biweekly by WebPosition Gold ([www.Webpositiongold.com](http://www.Webpositiongold.com)), software that tracked the search engine position of Web pages based on their keywords. Since some of the included search engines returned a list of only 200 ranked Web sites, engines where the URL of the experimental page did not appear or was ranked greater than 200 were recorded as having a ranking of 201. Baseline data on search engine position was collected from January 1999 to December 1999, prior to the implementation of the first design intervention.

## Design Intervention One

Between January 2000 and May 2000, two phases of design interventions were implemented to empirically test the value of the heuristics described above. As the Arthritis Source is a large Web site, a complete redesign to incorporate site-wide consistency (a characteristic theorized to be attractive to search engines) was out of the scope of the present experiment. Other heuristics, such as the length of each descriptive tag, did not follow the previously described heuristics precisely, as the compilation of heuristics was still in progress at the time of the experiment. Although the design interventions did not conform to our assembled heuristics exactly, they do, as the reader will see, follow them to a substantial degree.

(Insert Figure 2 here)

In the first experimental manipulation ([http://www.orthop.washington.edu/shoulder/Dislocat\\_1.htm](http://www.orthop.washington.edu/shoulder/Dislocat_1.htm), [RotCufSu\\_1.htm](http://www.orthop.washington.edu/shoulder/RotCufSu_1.htm), [ShoulRep\\_1.htm](http://www.orthop.washington.edu/shoulder/ShoulRep_1.htm)), the three test pages received new keyword and description meta tags to bring them into closer accordance with the keyword and consistency heuristics. The modified pages were then individually submitted to a selection of bot-based search engines. Nine new meta keywords, most of which were under 20 characters and repeated in the body of their associated page, were chosen to replace the existing meta tag keywords of the experimental pages. Hit frequency and keyword ranking data for the first design intervention were collected between January 1 and February 15, 2000.

## Design Intervention Two

In the second design intervention ([http://www.orthop.washington.edu/shoulder/Dislocat\\_2.htm](http://www.orthop.washington.edu/shoulder/Dislocat_2.htm), [RotCufSu\\_2.htm](http://www.orthop.washington.edu/shoulder/RotCufSu_2.htm), [ShoulRep\\_2.htm](http://www.orthop.washington.edu/shoulder/ShoulRep_2.htm)), the same three patient information pages were redesigned as a series of 7 or 8 shorter Web pages linked together into a linear sequence. Each of the three topics, in effect, became a small Web site, a subsection of "The UW Bone and Joint Sources." These shorter pages were each designed in accordance with the page design, meta tag, and consistency heuristics described above. Each page was limited to 250 words or fewer and included a title and topic sentence that followed the prescribed lengths. The keywords for the first page in each series did not change, and the secondary pages contained some of these primary keywords in addition to individual selections specific to each page.

(Insert Figure 3 here)

After the second iteration of design was complete, the pages were resubmitted to all of the search engines included in the first design intervention, as well as Yahoo!, Open Directory, and Snap, which are directory-based search engines. Hit frequency and keyword ranking data for the second design intervention was collected between March 1 and April 15, 2000.

## Results

In accordance with our first hypothesis, the aggregate search engine position of individual keywords, as seen in Figures 4, 5, and 6, reacted very differently to the two design interventions. The aggregate search engine position of 8 of the 27 keywords for all three pages were not affected by either of the design interventions. The other 21 keywords showed some improvement during the course of the experiment. Of those keywords whose aggregate rankings varied during the experimental period, most showed greater improvements in aggregate search engine position after the second design intervention than after the first. There were two exceptions to this trend: the keyword “glenohumeral instability” for Shoulder Dislocation Repair Surgery, and “acromioplasty” for Rotator Cuff Repair Surgery.<sup>4</sup> As predicted by our third hypothesis, increases in aggregate search engine position had a tendency to decrease with time. This effect is especially noticeable after the second design intervention: 15 of the 21 keywords that showed improvements had an increase in aggregate ranking shortly after Design Intervention Two (3/01 to 3/20), then began to decline around the beginning of April.

### ***A Metric For Evaluating Search Engine Position***

To ascertain changes in the overall ranking of each keyword over time, an aggregate search engine position was calculated as the sum of search engine position of each keyword across the 14 search engines included in the analysis. Thus, if ShoulRep.htm ranked 20 on Yahoo/Inktomi and 201 on HotBot for the search term “shoulder replacement”, these were summed as 221, and that sum was added to the ranking of the other engines. The best possible score using this scale would be 14 (1 x 14), which would indicate that the Web page was ranked first for all 14 search engines for a given keyword. The worst possible score for a given keyword would be 2814 (201 x 14), indicating that the experimental Web page associated with the search term was ranked above 200 or was not found on all 12 search engines<sup>5</sup>. This aggregate search engine position for each keyword was plotted against time to see how the ranked position of each keyword varied throughout the experiment. These results are presented in Figures 4, 5 and 6.

(Insert Figures 4, 5 and 6 here)

To ascertain the frequency of use of the three experimental pages over time, hit rate was plotted against time. These results are presented in Figures 7, 8, and 9. The linear correlation between the weekly average of hit frequency and weekly search engine position was then computed for the primary keyword for each of the experimental pages.<sup>6</sup> Lastly, the biweekly search engine position of all the keywords for all of the pages was summed within each of the search engines being tracked in the study. These sums were then plotted against time to measure the effectiveness of the design interventions on individual search engines.

Hit frequency to the three experimental pages, as seen in Figures 7, 8, and 9, did not drastically change after the first design intervention. However, a sharp rise in hit frequency occurred after the submission of the second design intervention<sup>7</sup>. This rise retained the high degree of day to day variability found in the baseline data. Each of the three pages had a different “peak” in daily hit frequency ranging from 37 (ShoulRep.htm) to 61 hits/day (RotCufSu.htm). These large initial increases in hit rate were lost over time and varied in duration from page to page. In the case of ShoulRep.htm (Figure 9), hit frequency spiked sharply in the third week after the second design intervention, only to drop to an average value closer to baseline. Dislocat.htm (Figure 7), on the other hand, peaked in the fifth week and kept a larger proportion of that increase throughout the remainder of the data collection period.

(Insert Figures 7, 8 and 9 here)

As predicted by our second hypothesis, the linear correlation between hit rate and the aggregate search engine position of the most representative keyword for each page showed an inverse relationship

<sup>4</sup> This is most likely due to the fact that these keywords appear on their respective pages only once.

<sup>5</sup> Our search engine position metric is not sensitive to rankings higher than 200.

<sup>6</sup> Only the most representative keyword for each page was included, as some of the keywords were less representative of the page's content and would not have been useful in this analysis.

<sup>7</sup> Spiders may have been returning to the site repeatedly during the data collection periods, inflating the hit frequencies. However, FastStats automatically removes bot & spider hits by identifying commonly used IP addresses and domain names. Most search engines release the name of their bot or spider so it may be identified in the log files.

between aggregate search engine position and hit rate. Dislocat.htm showed the strongest correlation ( $r = -0.848$ ), followed by RotCufSu.htm ( $r = -0.790$ ) and ShoulRep.htm ( $r = -0.728$ ).

## Discussion

We maintain that learners who are searching for educational materials on the Web (outside of those assigned in a curriculum) often use search engines to find information for their learning tasks. Given that bot-based search engines use predictable algorithms to rank Web sites, we have hypothesized that educators can promote the rankings of their Web-based educational materials by designing them to be attractive to those algorithms and by submitting their URLs to search engines. Since these efforts at promoting educational Web sites can improve the visibility of those Web sites, we have hypothesized that more learners will be able to find and visit them. The results of the test of our design heuristics suggest that the rankings of Web pages based on their keywords can be improved with design. As a result, these improvements in search engine position are correlated with increased hits.

The search engine position data from the first design intervention suggests educators can make improvements in the rankings of their Web sites simply by changing the site's metadata and submitting it to a variety of engines. While this process follows the commonly suggested methods of improving rankings, it was not as effective as the more intensive process used in the second design intervention. The greater degree of improvement during the second design intervention suggests that the creation of groups of related pages that are short in length (250 words in our case) and have internally consistent meta tags and content is a better method of optimizing search engine position.

According to the theory embodied in the consistency heuristics, this effect may have been more pronounced if the rest of the Arthritis Source Web site was designed in a similar fashion.

The decline of search engine position over time illustrates the effects of sustained competition for rankings. We maintain that this loss in ranking is due to the creation or resubmission of other sites on the same topic. That this trend was seen in our data suggests that periodic resubmission is essential to the maintenance of search engine position. Based on the data in this study, we recommend monthly resubmission in order to sustain rankings.<sup>8</sup>

Some of the keywords for each page proved to be less representative of the content. The rankings of these terms tended to be less responsive to the design interventions. For example, the "glenoid" keyword for ShoulRep.htm was not representative of the content of its page. Figure 3 illustrates that the search engine position for this keyword did not change after the design interventions. These keywords did not show the same degree of improvement as other keywords because they were less relevant to the text of their associated pages. This effect suggests the importance of appropriate keyword choice in the creation of meta tags.

## Conclusion

We conclude by suggesting that the present experiment is preliminary evidence that the rankings of educational Web sites can be optimized in bot-based search engines. Search engine positioning strategies are always changing (<http://searchenginewatch.com>), and further research on optimization should explore the effects of newer algorithms, such as those that track the number of referring links or return visitors to a Web site.

Educators who are interested in maintaining highly ranked sites may need to become experimenters themselves, seeing how different design features and metadata affect the search engine position of their materials. Increasingly sophisticated software packages and commercial services are making these sorts of experiments relatively easy to conduct. However, it is important to note that this effort toward promoting a site may be better spent in other areas, especially if applying the design heuristics would compromise the educational utility of a Web site. A referring link from a heavily trafficked site on a subject may prove to be a better means of reaching learners-at-large than search engine optimization. Lastly, it is important to note that search engine optimization may not pose a challenge for some educators, as the subject of their learning materials may be relatively unique and free from strong ranking competition. In these cases it remains important to optimize keywords and metadata to the search behavior of intended users of these learning materials.

---

<sup>8</sup> However, some search engines discourage overzealous resubmission of Web sites by categorizing such sites as spam. WebPosition recommends submitting a maximum of once per week, although once per month appears to be sufficient.

The relevance of search engine position for educators may become less significant several years from now. As educational resource gateway web sites become widely known and utilized by learners-at-large, it may become more useful for educators to register with these portal sites. As gateway sites gain prominence, they hold the promise of promoting peer review of web-based educational resources and raising the quality of these resources. Traditional search engines have access to less than one percent of the total documents and media resources available on the web, but new specialized types of search engines are now available to learners-at-large (New York Times, 2001). These new types of search engines may be best suited to locating multimedia resources and resources in specialty areas such as medicine and patient education.

While the optimization of search engine position can play an important part in extending the mission of educational institutions, locating a credible Web site with appropriate material is only the first step. Even if a learner “hits” a page, they have not necessarily learned anything from its content (the primary goal of the educator who creates an educational site). Although fields such as patient informatics have begun to explore the impact of learning-at-large, investigation into this type of learning will continue to gain importance as the issue of access to learning materials becomes increasingly critical.

## References

- Bader, S. A., & Braude, R. M. (1998). "Patient informatics": creating new partnerships in medical decision making. Academic Medicine, 73(4), 408-11.
- Barab, S. A., Bowdish, B. E., & Lawless, K. A. (1997). Hypermedia navigation: Profiles of hypermedia users. . Educational Technology Research & Development, 45(3), 23-41.
- Beredjiklian, P. K., Bozentka, D. J., Steinberg, D. R., & Bernstein, J. (2000). Evaluating the source and content of orthopaedic information on the Internet. Journal of Bone & Joint Surgery, 82A(11), 1540-1543.
- Chi-Lum, B. (1999). Friend or foe? Consumers using the Internet for medical information. Journal of Medical Practice Management, 14(4), 196-8.
- Dyer, K. A., Thompson, C. D., Reis, O., & Romer, S. (1998). Using the Internet for Patient and Physician Web-Education and Health Promotion. World Congress for the Internet in Medicine, Virtual Congress, MEDNET '98, London, England.
- Guernsey, L. (2001, January 25). Mining the 'Deep Web' with specialized drills. New York Times on the Web. Retrieved January 25, 2001, from the World Wide Web: <http://www.nytimes.com/2001/01/25/technology/25Sear.html>
- Hardey, M. (1999). Doctor in the house: the Internet as a source of lay health knowledge and the challenge to expertise. Sociology of Health & Illness, 21(15), 820.
- Hill, J. R. (1999). A conceptual framework for understanding information seeking in open-ended information systems. Educational Technology Research & Development, 47(1), 5-27.
- Kinze, M. B., Larsen, V. A., Burch, J. B., & Boker, S. M. (1996). Frog dissection on the world-wide Web: implications for widespread delivery of instruction. Educational Technology Research & Development, 44(2), 59-69.
- Lawrence, S., & Giles, C. L. (1998). Searching the World Wide Web. Science, 280(5360), 98-100.
- Leaffer, T., & Gonda, B. (2000). The Internet: an underutilized tool in patient education. Computers in Nursing, 18(1), 47-52.
- Lorig, K., Konkol, L., & Gonzalez, V. (1987). Arthritis Patient education: a review of the literature. Patient Education & Counseling, 10, 207-52.
- McCullough, M. (2000, April 24). Virtual medicine: promise and peril. Philadelphia Inquirer, p. D1.
- Nogler, M., Wimmer, C., Mayr, E., & Ofner, D. (1999). The efficacy of using search engines in procuring information about orthopedic foot and ankle problems from the World Wide Web. Foot & Ankle International, 20(5), 322-5.
- Norman, D. A., & Spohrer, J. C. (1996). Learner-centered education, Communications of the ACM, 39(4), 24-27.
- Saccetti, P., Zvara, P., & Plante, M. K. (1999). The Internet and patient education-resources and their reliability: focus on a select urologic topic. Urology, 53(6), 1117-20.
- Seedhouse, Paul. (1999). Task-based Interaction. English Language Teaching Journal, 53(3), 149-156.
- Shneiderman, B. (1998). Relate-Create-Donate: a teaching learning philosophy for the cyber-generation. Computers in Education, 31, 25-39.
- Soloway E., & Pryor A., (1996). The next generation in human computer interaction. Communications of the ACM, 39(4), 16-18.
- Soot, L., Edwards, J. M., & Moneta, G. L. (1999). Vascular Surgery and the Internet. Journal of Vascular Surgery, 30, 84-91.
- Starr, R. M. (1997) Delivering Instruction on the World Wide Web: Overview and Basic Design Principles. Educational Technology, 37(3), 7-15.
- Tauscher, L., & Greenberg, S. (1997). How people revisit Webpages: empirical findings and implications for the design of history systems. International Journal of Human-Computer Studies, 47, 97-137.
- Tunender, H., & Ervin, J. (1998). How to succeed in promoting your Web site: the impact of search engine registration on retrieval of a World Wide Web site. Information Technology & Libraries, 17(3), 173-179.
- Whittington, C. D., & Campbell, L. M. (1998) Task-Based Learning Environments in a Virtual University. Computer Networks & ISDN Systems, 30, 707-709.

## Tables and Figures

**Table 1. Attributes Visible To Common Robots And Crawlers**

Keywords meta tags	<meta name="keywords" content="keywords search terms descriptive words">
Description meta tag	<meta name="description" content="This descriptive sentence should summarize the contents of the page.">
Title tag	<title>Title of the page</title>
Body text	<body>Any text that appears here...</body>
Image tag alt attribute	
Noframes text	<noframes>Any text that appears here...</noframes>
Crawler pages	<meta name="robots" content="noindex, follow"> (Any text in http://your.domain.com/robots.txt)

**Table 2. Search Engines For Which Rankings Were Tracked**

Altavista	Google	Infoseek	MSN	WebCrawler
AOL	Goto	Lycos	Northern Light	Yahoo
Excite	Hotbot	Magellan	Snap	

**Figure 1. Dislocat.Htm Before Design Intervention**

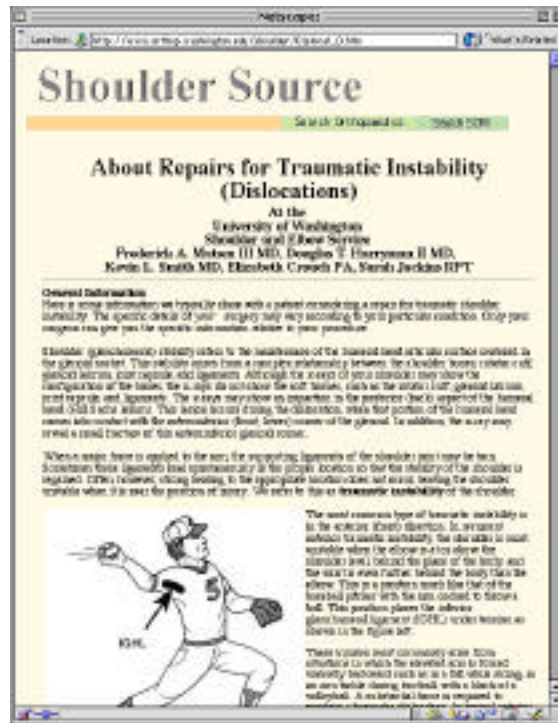


Figure 2. Dislocat.Htm After Design Intervention 1

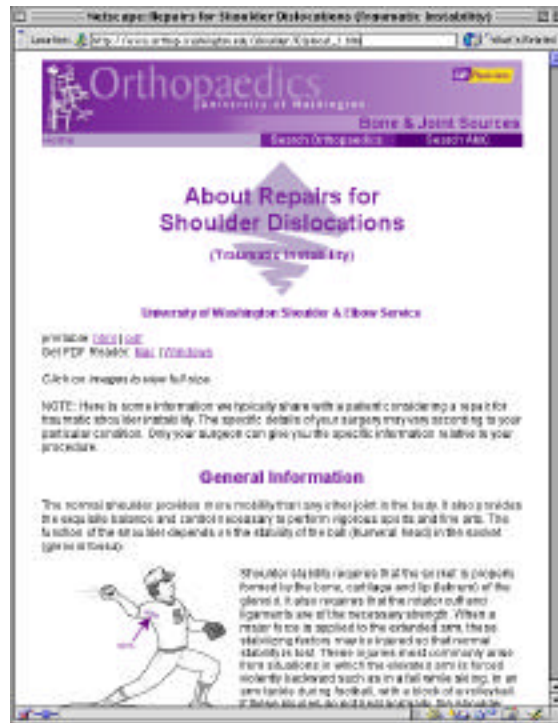
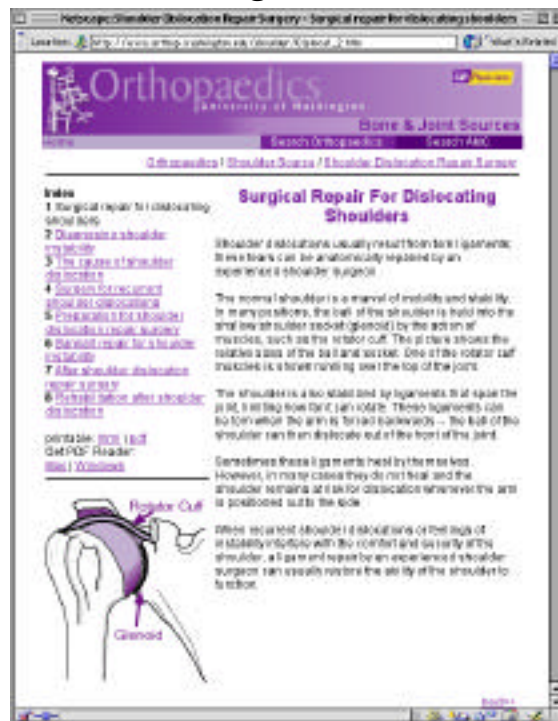
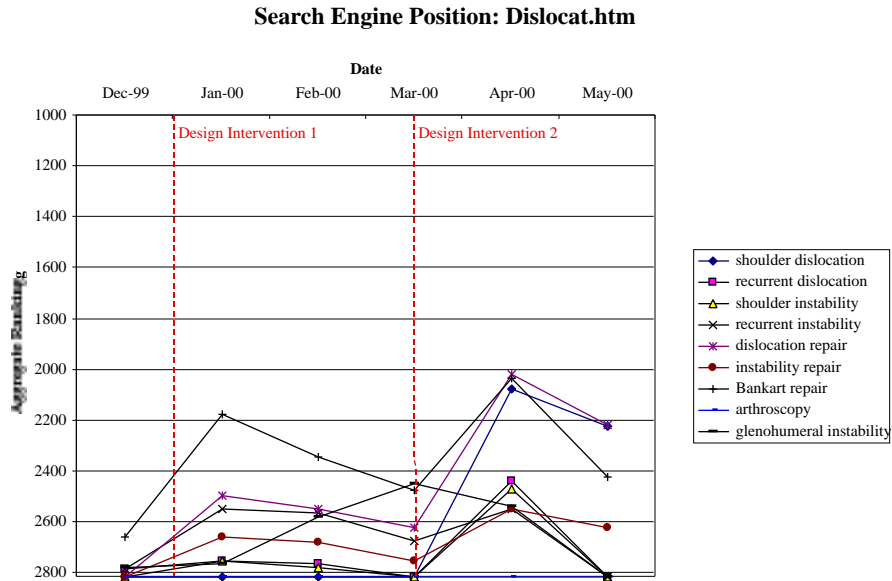


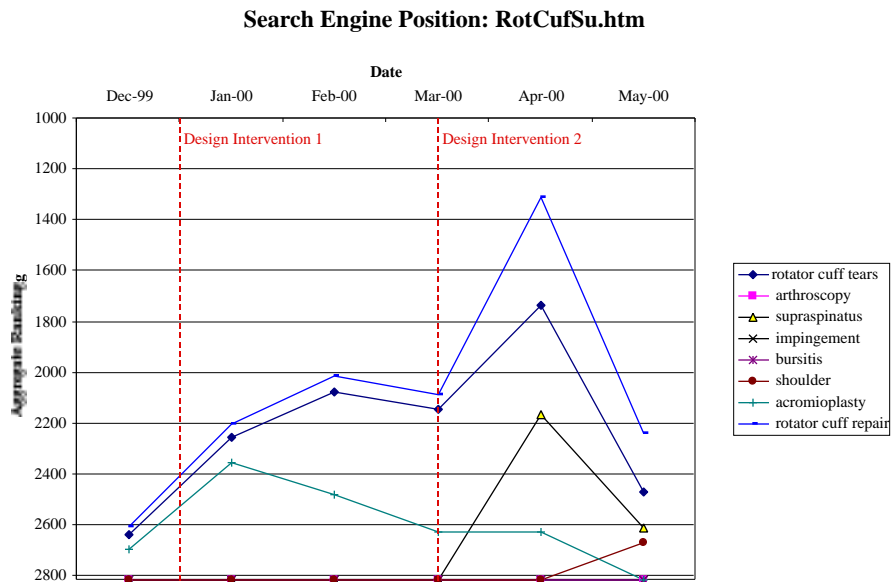
Figure 3. Dislocat.Htm After Design Intervention 2



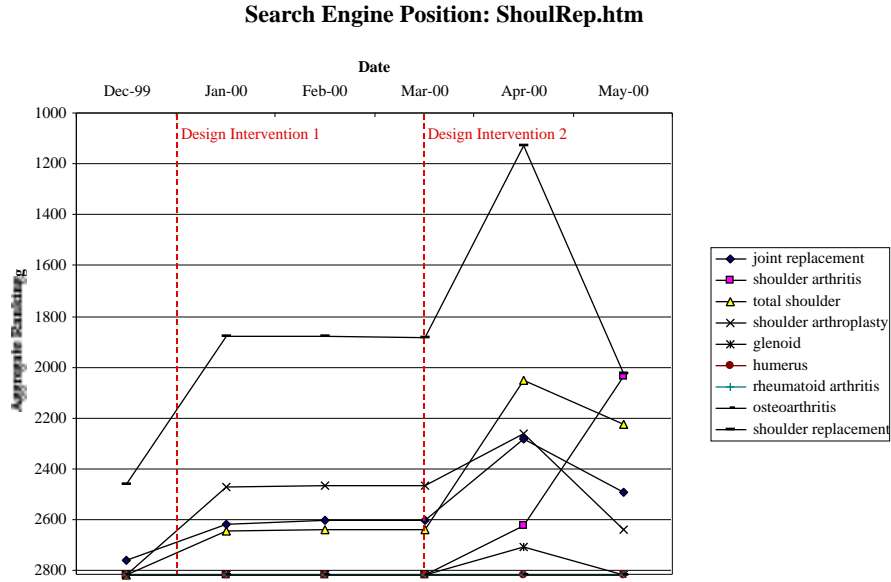
**Figure 4. Search Engine Position: Dislocat.htm**



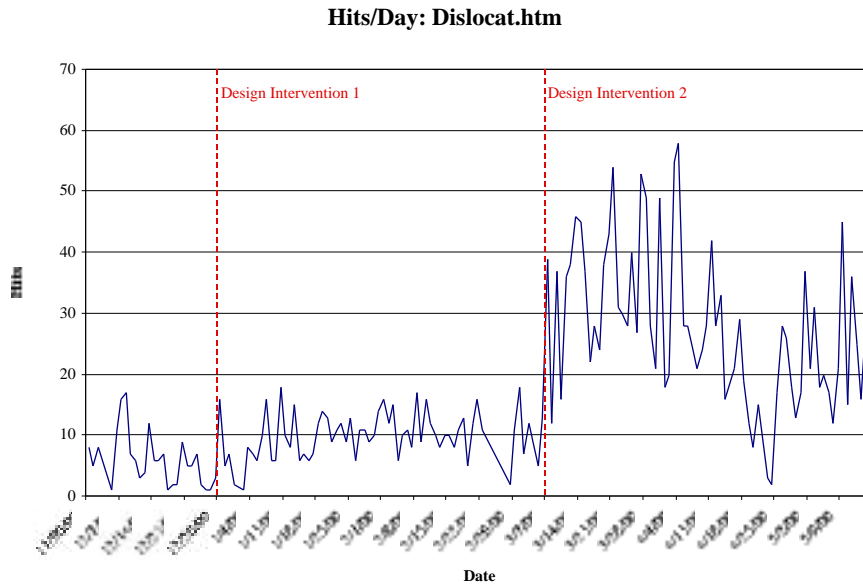
**Figure 5. Search Engine Position: RotCufSu.htm**



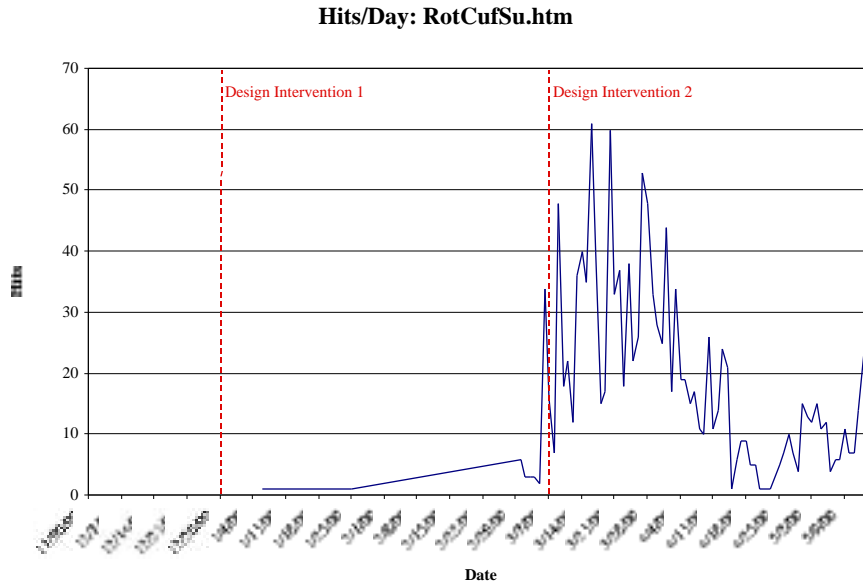
**Figure 6. Search Engine Position: ShoulRep.htm**



**Figure 7. Hits/Day: Dislocat.htm**



**Figure 8. Hits/Day: RotCufSu.htm**



**Figure 9. Hits/Day: ShoulRep.htm**

