DBLP Bibliography Website Final Report and Analysis

Members:  Ryan Flynn
          Patrick Brennan
          Christopher Bair
          Brian Watkins
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>5</td>
</tr>
<tr>
<td><strong>PART I - INTRODUCTION</strong></td>
<td>3</td>
</tr>
<tr>
<td>Background on the website and the author</td>
<td>3</td>
</tr>
<tr>
<td>Website goals and objectives</td>
<td>3</td>
</tr>
<tr>
<td>Importance of analyzing the website</td>
<td>3</td>
</tr>
<tr>
<td>User analysis and the traffic patterns of the website</td>
<td>4</td>
</tr>
<tr>
<td><strong>PART II – PERCEPTION ANALYSIS</strong></td>
<td>5</td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
</tr>
<tr>
<td>Methodology</td>
<td>5</td>
</tr>
<tr>
<td>Analysis</td>
<td>6</td>
</tr>
<tr>
<td><strong>PART III - SEARCH ANALYSIS</strong></td>
<td>8</td>
</tr>
<tr>
<td>Introduction</td>
<td>8</td>
</tr>
<tr>
<td>Screenshots</td>
<td>8</td>
</tr>
<tr>
<td>Methodology</td>
<td>11</td>
</tr>
<tr>
<td>Analysis</td>
<td>12</td>
</tr>
<tr>
<td><strong>PART IV - ALTERED PERCEPTION ANALYSIS</strong></td>
<td>19</td>
</tr>
<tr>
<td>Introduction</td>
<td>19</td>
</tr>
<tr>
<td>Methodology</td>
<td>20</td>
</tr>
<tr>
<td>Analysis</td>
<td>21</td>
</tr>
<tr>
<td><strong>PART V - POTENTIAL FOR SOCIAL NETWORKING</strong></td>
<td>22</td>
</tr>
<tr>
<td><strong>PART VI - CONCLUSION</strong></td>
<td>24</td>
</tr>
<tr>
<td><strong>PART VIII - CITATIONS</strong></td>
<td>26</td>
</tr>
<tr>
<td><strong>PART IX - APPENDIX A</strong></td>
<td>27</td>
</tr>
<tr>
<td><strong>PART X - APPENDIX B</strong></td>
<td>28</td>
</tr>
</tbody>
</table>
Abstract

Our team has analyzed the DBLP website with specific focus on how well a user is able navigate through the website to find what they are looking for. We chose this website because it is a valuable resource for finding bibliographical information on a wide variety of publications including computer science articles. The website is frequently used by students and academic professors. Our group focused on how well the text of the website draws the user’s attention to important site functions along with the three individual search functions that are listed on the homepage. Our goal is to provide suggestions of ways to alter the website to maximize utilization of the proper search functions in order to facilitate faster more effective browsing of the stored information.
Part I- Introduction

Background on the Website and the Author

DBLP, commonly known as Digital Bibliography and Library Project, is a website that acts as a computer science bibliography. It originally was a database and logic programming bibliography site created in the early 1980s. The website is hosted at Universitat Trier in Germany. In March 2008 the site listed over one million articles related to computer science.

Michael Ley is the creator, author, and holds sole responsibility for maintaining the site. He received an award from the Association for Computing Machinery in 1997 for his work. He is also a lecturer at the Universitat Trier in Germany.

Website Goals and Objectives

The website is used to find articles, journals, and other works dealing with computer science. It is a massive database and archive of works from many authors. You can search through this database in various ways and search functions. Overall the site provides a massive database for users to search through different computer science and technology related sources.

Importance of Analyzing the Website

We wanted to analyze the DBLP site because we found it very interesting and an important tool to use for many different types of users. When we first saw the site, we saw potential ways we could make it more efficient and aesthetically pleasing to make a better experience for the user. It is very important for a site like DBLP to maintain maximum usability so it is easy for the visitors to finish the task at hand in the most efficient way. In a very busy world we live in, time is money. Also websites must be designed toward the right audience. The actual users of the
site are very important in the designing of the interface. We feel that because this site is such an important tool, we want to make it the best possible environment for the user.

**User Analysis and the Traffic Patterns of the Website**

To better understand the goals and objectives of the DBLP Website, it is necessary to analyze and understand the users of the website. Analyzing the website we were able to divide users into three primary categories: Students, Academic Professors and Instructors, and Researchers and Scientists. Each category has differing goals and objectives. Although all users cannot be generalized into these specific categories, by defining the most important users, we will be able to better understand the most important aspects of the website.

**Students**

Users from this category can include Undergraduate, Graduate, and Doctrinal students seeking information and citations for class projects, papers, and research. This group is highly diverse and includes users with many different levels of academic understanding and motivation, as well as different goals and objectives. Undergraduate students are using the website with limited knowledge and subject matter understanding as well as limited social networks and connections with collaborators and authors on the website. On the other hand, doctrinal students have a greater use and understanding of the website as well as a larger social connection with collaborators and authors.

**Academic Professors and Instructors**

Academic professors and instructors provide a large percent of those that collaborate and author many articles that are on the website as well as using the site to look up information on colleagues work and other articles that can be used for classroom resources, citations for
assignments and papers, as well as course material. These users are highly educated, up-to-date with computer science and technology and have a large social network with other users and collaborators on the website.

**Researchers and Scientists**

Researchers and Scientists also make up a large percentage of both users to the DBLP website and authors to many journals and articles listed on the site. These users often times are on the cutting edge of research for computer science and use the site to obtain information about topics of interest and similar research to their own. These users and very highly educated and create many of the advances and milestones in computer science and have a large social network and collaboration base of many other researchers, scientists, and professors in the field.

**Part II – Perception Analysis**

**Introduction**

This experiment was to determine how well the size, boldness, and the color of the text would catch the user’s attention. A page from the website, the International Conference on Conceptual Modeling, was modified to display different size text. The subject was then asked as series of questions on content of the bibliography. Subjects were timed for comparison between the various sized text and the normal sized.

**Methodology**

This task was based on a previous lab that studied how a user perceives certain types of text and how well the differently altered text would attract the attention of the user. The International Conference on Conceptual Modeling bibliography web page from the DBLP website was modified to have certain areas of text with different sizes while other areas were rendered bold.
Dates, titles, and ISBN numbers were some of the text segments that were altered for the experiment.

The subjects were asked a series of ten questions and were timed for each individual question. The time continued to run until they found and reported the correct answer. After each question, the subject would return to the default starting point at the top of the page. The answers to half of the questions were of different text sizes while the other half were left unaltered as a control group. The results were then analyzed to determine the presence of any form of trend.

**Analysis**

The performance of each of our subjects is depicted graphically in figure 2.1. The text the subjects were required to use in order to answer each of the ten questions had been altered in such a way to encourage or discourage a quick response. The questions that were asked can be found in appendix A. Questions 2, 3, 6, and 9 had certain elements within the relevant entry enlarged by varying degrees. Question 7 had a certain element within the relevant entry decreased in size. The intended result was to record lower response times for questions 2, 3, 6, and 9 as well as to record question 7 as having the longest response time.
The results obtained from this experiment were somewhat on track with our group’s hypothesis however question 3 caused a major hiccup in our data. The response times for question 3 are an almost exact match to the response times for question 7. This is a significant problem seeing as question 7 had elements of its entry significantly decreased in size while question 3 had elements of its entry increased in to a much larger size than any other element. This should not have happened but, aside from the data of subject 4, who had consistent results across almost the entire series, along with subject 5, who exhibited times so far outside the overall trend that we omitted his results, question 3 boasts some of the longest response times.
Part III – Search Analysis

Introduction

An experiment was created to determine the effectiveness of the search engine, as well as how
the displayed results make it easier for the user to find what they were searching for. There are
three different search engines in the DBLP: an author search engine, a faceted search engine, and
a complete search engine. We asked our subjects, Penn State University students, questions that
had to be answered using each of the search engines while also timing them. The questions asked
can be found in Appendix B. We used these results to determine how each search engine
performed for the questions asked.

Screenshots of Three Search Engine Functions

[Figure 3.1- Subject Author Search Function.]
[Figure 3.2- Subject Author Search Results Page]

[Figure 3.3 - Faceted Search Function.]
Figure 3.4 - Faceted Search Results Page.
Methodology

The purpose of this task was to test the functionality of the three search engines that are provided on the DBLP website. These search engines include “Author”, “Faceted”, and “Complete”. To analyze the effectiveness of each search, we asked the subject ten questions to be answered with each of the search engines that can, or can not, be answered with the individual search. The reason for this is to point out how some of the search engines are ill-equipped for finding certain types of information. The subjects were timed from when they started typing in the search bar, starting from the original search page. If the subject could not find the answer and gave up, a
note was made in the analysis about the usefulness of the particular search compared to the question. The times were compared to find the most practical search engine for our website.

**Analysis**

The DBLP website has three different search functions available on its main page to facilitate browsing of its archives. These are “Author”, “Faceted Search”, and “Complete Search”. Our group discovered an interesting trend while examining a past lab. Despite having been given no guidelines as to which search function(s) they were allowed to use, each and every subject tested choose to use only the search function labeled “Author”. We found this to be worth further investigation so we experimented with the other two search functions and after some time we developed the hypothesis that the “Author” search function was, by far, the least useful out of the three search functions. A range of questions were designed to test the user compatibility and general competence of each search function. Each question incorporated one or more constraints including author(s), date(s), topic, number of publications, etc. Some of the questions were designed to be unsolvable when utilizing the simpler search functions.

Figures 3.6 through 3.12 show the response times for subjects A through G with respect to the three search functions. Working under our hypothesis it was believed that each subject would produce results that would follow two basic trends: (1) Response times would be longest and total answers fewest with respect to the “Author” search. (2) Response times would be shortest and all questions would be answered when utilizing the “Complete Search”. The “Faceted Search” was expected to produce sporadic results and was not expected to answer every question. Trend 1 was partially adhered to in all cases however that was never a real issue as the “Author Search” was designed to fail when presented with sufficiently complicated questions.
Unfortunately, our analysis indicates that in the instance where the user intends to find data pertaining only to an author, with no other constraints, the “Author Search” is often the faster choice. Trend 2 was best adhered to with regards to Subject D (Figure 3.9). Subject D meets all the criteria of our hypothesis save one. The subject was unable to answer question 5 using any of the three search functions. Another anomalous behavior of interest was exhibited by subject 4. Subject 4 was the only subject to attempt to move into another of the three search functions when the “Author” search proved unsuccessful. He was instructed that this was not allowed however it is interesting that he was the only one to consider this option seeing as none of the subjects were explicitly told not to use the other functions.

We now believe that when it comes to searching for data, personal preference plays a significant role. The large variation in times can be attributed to the varied ways different users approach a problem. In a multifaceted query a user may chose to begin their search with any one of the different aspect of the query or they may enter everything at once and see if they get lucky. While this may be the case most of our subjects chose to begin their search with a single element and narrow the parameters as they went. By disallowing the use of built in browser search functions we limited the users abilities to sift though an overwhelming amount of text. Certain users were prone to having the necessary information displayed on the screen but being unable to see it.
Figure 3.6 – Subject A’s response time data for the Search Lab.

Figure 3.7 – Subject B’s response time data for the Search Lab.
Time Between Question and Response (sec)

[Figure 3.8 – Subject C’s response time data for the Search Lab.]

Time Between Question and Response (sec)

[Figure 3.9 – Subject D’s response time data for the Search Lab.]
Figure 3.10 – Subject E’s response time data for the Search Lab.

Figure 3.11 – Subject F’s response time data for the Search Lab.
The large variation in times can be attributed to the varied ways different users approach a problem. In a multifaceted query a user may chose to begin their search with any one of the different aspect of the query or they may enter everything at once and see if they get lucky. While this may be the case most of our subjects chose to begin their search with a single element and narrow the parameters as they went. By disallowing the use of built in browser search functions, we limited the users abilities to sift though an overwhelming amount of text. Certain users were prone to having the necessary information displayed on the screen but being unable to see it. Also relevant to this analysis is the fact that our group utilized users with no prior experience operating the DBLP website. The wide range of behavior observed in this lab can be attributed to the fact that we did not utilize experts for the process. While expert behavior can be predicted with relative accuracy, the actions taken by amateurs vary significantly given static conditions and thus can be highly difficult to anticipate.
[Figure 3.13 – Subject response time with respect to the posed question using “Author” Search]
[Figure 3.15 – Subject response time with respect to the posed question using “Complete Search”]

Part IV – Altered Perception Analysis

Introduction

This experiment was designed after the analysis of all previous labs had been completed. Once we determined that increasing the size or the level of boldness of an area of text would cause a user’s attention to be drawn more quickly towards it, and we had determined that the “Complete Search” function was the most effective, though not always the most efficient, search function available on the DBLP website, we set about designing a lab to determine if we could increase the number of users who would be drawn into using the “Complete Search” function. Our group redesigned the DBLP main page to attempt to elicit this result.
**Methodology**

As stated beforehand, when our group asked a group of users, in a past lab, to search for the coauthor of a particular publication, given only the authors name, every one of them choose to search using the “Author” search function. Through manipulation of simple html elements we bolded the “Complete Search” function and moved it into the position previously held by the “Author” function. We then proceeded to run a new group of users through same process, posing the same question, as was used in the old lab, however this time we recorded not the time it took them to find an answer but simply which function they used.

![Figure 4.1 – Altered DBLP Main Page with Altered Section Boxed in Red](image)

**The DBLP Computer Science Bibliography**

maintained by Michael Ley - Welcome - FAQ

DBLP is available from several hosts: Trier I - Trier II - ACM SIGMOD - VLDB - SunSITE DE

**Search**

- **CompleteSearch** (Trier East, Max Planck Institute for Inf)
- **Author**
- **Faceted Search** (LIS Research Center, U. Hannover)

**Bibliographies**

- **Conferences**: SIGMOD, VLDB, PODS, EE, ICDE, POPL, ...
- **Journals**: CACM, TOCS, TOPLAS, PVLDB, Inf. Systems, TPLF, TCS, ...
- **Server**: LCS/INRIA, IPP
- **Books**: Reference - Collections - DB Textbooks
- **By Subject Database Systems, Logic Prog, Tk, ...

**Full Text**: ACM SIGMOD Anthology

**Links**

- **Computer Science Organization**: ACM (DL / SIGMOD / SIGIR), IEEE Computer Society, DL, IEEE Xplore, IJIP
- **Related Services**: Google Scholar, Citeseer, CrossRef, DBLP, dblp, eprint.net, CoRR, HAL, ICDE, Zentralblatt Math, MathSciNet, BibSonomy, CiteULike, Libra, ArXiv, ...

[Figure 4.1 – Altered DBLP Main Page with Altered Section Boxed in Red]
A total of 13 individuals were tested on the altered website. Out of those 13 subjects 11 choose to use the “Complete Search” function and 2 choose to use the “Author” search function. None of the subjects choose to use the “Faceted Search” function. Figure 4.2 shows the function usage breakdown by percentage.

![Figure 4.2 – Search Function Usage by Percentage](image)

The previous search function usage breakdown, observed under the conditions of the original, unaltered main page, was 100% “Author” search function and 0% usage by the other two. This lab has proven that by bolding the “Complete Search” function, and relocating it to the top of the search function list, its visibility, and subsequent usage by amateur users, has been substantially increased/improved. By making these, or similar, adjustments to the actual DBLP website, users would be more likely to gain quick, easy access to the information they are seeking.

Part V – Potential for Social Networking
Towards the end of the 20th Century, the Internet was just beginning to move into a position of providing data and information to people in mass quantities; a time referred to as the dawn of the Information Age. Today, the Internet has grown and expanded into a platform that not only supplies the world with information, but also allows for endless networking opportunities with countless people. This movement from the World Wide Web has been coined Web 2.0 and is changing the way people connect, communicate, and learn. According to Tim O’Reilly, CEO of O’Reilly Media and author of the popular blog O’Reilly Radar, defined Web 2.0 as, “the business revolution in the computer industry caused by the move to the internet as a platform.” This platform is seen as the ability for users to connect to users from around the world that are pursuing the same goals, objectives, and interests. This connection allows for the exchange of data and for communication and a level never before experienced. Researchers are able to share data, results from experiments, and conclusions that have been reached. This allows for collaboration between colleagues, not only in the same location, but thousands of miles away. This connection is bridging the gap of communities and cultures and is allowing for the globalization of information and ideas.

The DBLP Website provides users with information relating to the authors, co-authors, and collaborators of thousands of papers, journal articles, periodicals, and other writings. This massive database of scientific articles attracts many academic users, researchers, and scientists to view information provided by other peers and colleagues on similar computer science and technology related concepts. However, the website does not allow for contact between users, authors, and collaborators to further discuss research information or ideas. Due to the current information provided, the DBLP website has potential to create social networking aspects to allow more contact and communication between users. Providing information for each author
to be contacted through office or email information could allow for further collaboration between users. This potential for contact could allow for partnerships, sharing of research information as well as results, helping to bring together the scientific community. Information could also be linked to authors’ websites and research websites to provide users with more information.

Another aspect of Web 2.0 technology that could be applied to the DBLP site is to allow for user generated feedback to be posted on articles and authors for users to view. This Web 2.0 feature is best seen on the website Linked-In, a professional social networking site for businesses, which allows users to post comments about colleagues and create feedback as to the merit and work ethic of individuals. This feedback has proven pivotal to potential employers and colleagues to be able to view feedback on possibly working with users in the future. Including this feature on the DBLP Website would allow users to leave feedback about articles as well as authors that can be viewed by other users. DBLP could use a similar feedback system to Linked-In or even eBay, so that users could obtain a better view of the work completed by authors and collaborators on the website.

Applying social networking aspects to the DBLP Website would allow for the site to evolve from just providing information about authors and articles to a platform of providing knowledge, ideas, and collaboration from users from many different backgrounds and disciplines from around the world.
Part VI – Conclusion

I believe that what we have all learned from this experience is that no matter how well a product functions if it lacks an interface which allows full utilization of its functionality towards its intended purpose, it will be received poorly by users. First impressions of the DBLP website lead our group to believe that it was an inferior piece of work perpetrated by a sub par web developer. Only after being forced to interact with the site for an extended period did we become aware of its impressive nature. The usefulness and overall professional construction of the site were masked by its uniform, lackluster appearance. If I had been searching for bibliographic information for a paper, or other project, I would have taken one look at this site and closed the window. We understand perfectly well that this site is intended for professional level users however there is nothing wrong with attempting to broaden your user base.

Important functions need to be easily identified and should catch the user’s eye whether they were attempting to locate that function or not. Users don’t usually want to search everything in front of them before they find what they want. They want what they are looking for to find them. The most useful functions should be located near the top of the page, or at the top of a list of available options, and should be made more visible by increasing the size of the text or using a bold font.

Unnecessary or redundant features, especially those that possess inferior functionality, should be removed in order to avoid undue stress and confusion on the part of the user. Users may begin using one particulate function and be unaware that a more useful function even exists. If the function in use proves ineffective the user may simply cease interaction altogether. It is of great importance to insure that a user is made aware of the most effective method to complete a task or
is made aware of methods to complete a task. As it is difficult to ensure the later with a high degree of certainty we recommend ensuring the former.


1.) What were the dates when the conference was held in Barcelona, Spain?

2.) What were the dates when the conference was held in Tuscan, Arizona?

3.) What year was the conference held in Manchester, U.K.?

4.) What is the ISBN number for the conference held in Toronto, Canada?

5.) What were the dates when the conference was held in Chicago, Illinois?

6.) Where was the conference October 9th - 12th, 2000?

7.) What year was the conference held in Lausanne, Switzerland?

8.) What were the dates the conference was held in Tampere, Finland?

9.) What is the ISBN number for the conference held in Los Angeles?

10.) What is the ISBN number of the conference held in Cottbus, Germany?
1. How many books did Paul L. Rosin author with Geoff A. W. West?

2. What year did Gary Feldman author a book with Ralph L. London?

3. How many books did John Mylopoulos author in 2005?

4. What year, within the past decade, were the most books on neurocomputing published?

5. Who published more than 100 books in a single subject at the turn of the millennium?

6. How many books have no listed date of publication?

7. What two authors have published the most books on piracy?

8. When was the first book, listed on the site, regarding Artificial Intelligence?

9. Who wrote the most recent book regarding R.E.M. sleep?

10. How many authors have written books with Shan Yan?