

Augmenting the Exploration of Digital Libraries with Web-Based Visualizations

Peter Bergström and Darren C. Atkinson
Department of Computer Engineering, Santa Clara University
500 El Camino Real, Santa Clara, CA 95053
{pbergstrom, datkinson}@scu.edu

Abstract

Web-based digital libraries have sped up the process that scholars use to find new, important research papers. Unfortunately, current digital libraries are limited by their inadequate webpage-based paradigm, and it is easy for even the most experienced scholar to get lost. A paper and its immediate references are shown on a webpage, but it is not obvious where that paper belongs in the larger context of a field of research. The goal for our research was to develop and test the effectiveness of a web-based application, PaperCube, that was designed to augment a scholar's interaction with a digital library and explore bibliographic meta data using a defined set of visualizations. These visualizations needed to provide different levels of visibility into a paper's citation network without losing focus of the currently viewed paper. PaperCube was validated through a user study which showed that it was very useful when it comes to augmenting digital library search by reducing the "cognitive load" put on a scholar and aiding the "discoverability" of new research material.

1. Introduction

When a scholar develops new concepts, ideas, and inventions, he or she most often builds upon, and expands beyond, previously published research. Therefore, a scholar needs to find relevant research material, most commonly in the form of articles, papers, or books, and then forage for other works through references. Even if the person is an experienced and established scholar in a field of research, finding relevant material can be a difficult and time-consuming task. Even worse, for a newcomer to a field, without proper guidance, finding research material can be near impossible.

Furthermore, the rate at which scholarly papers are published is accelerating. The National Science Foundation [13] has shown that the world-wide publication rate of science and engineering research has grown at a steady annual rate of 2.3% between 1995 and 2005. Also, between 1988 and 2005, the amount of cross-institutional author collabora-

tion grew from 40% to 61%. Since the amount of published research material is growing and the interconnectedness of author collaboration is becoming increasingly tangled, it is important to create new methods to aid the discovery of relevant research.

1.1 Digital Libraries

Even before advent of the web, digital libraries in other forms were available either as stand-alone systems or through proprietary means via the early Internet. Services [10] were developed beginning in the 1960s. However, these services were specialized and, due to the nature of the technology at the time, inaccessible to most. Once the web started to be widely used in the mid 1990s, the way that information and knowledge was disseminated changed drastically, not only for digital libraries, but in general. Digital libraries have taken full advantage of the web to make scholarly research easy to access.

Unfortunately, although the web has made it easy to access publications, it has not solved one major problem: with the amount of knowledge available, sensory overload is a very real possibility. Due to this, the actual information a scholar is searching for can be lost in the clutter of irrelevant material. Searching for papers in a digital library (DL) makes it easy to narrow down the search space quite effectively, but the webpage-based user interfaces currently in use are not always adequate. The current paradigm of webpage-based searching, viewing, and foraging has its limitations; while focused on a paper, represented as a webpage in the DL with links to its referenced papers, it is non-trivial for a user to see where the paper belongs within the wider context. Without navigating away from the original paper, it is not possible to view the references of a referenced paper, and so on, or in other words, the overall citation network. In order to view a reference, a researcher has to click on a link that either makes the browser navigate away from the current page or opens a new browser window. Neither of these solutions are ideal because the researcher has left the original paper behind.

Therefore, it is important to allow for the navigation of bibliographic meta data in such a manner that allows a paper to be focused yet allow for the easy and seamless access to its surrounding citation network without losing focus and context. To create a practical system that allows for this and is web-based would be quite useful to scholars. Such a system does not necessarily need to replace currently existing web-based digital libraries, but rather augment them. Simply by replacing the webpage that represents a paper within the typical DL of today with a rich, interactive visualization-based user interface was thought to be a compelling challenge. However, it is important to remain web-based so that it is possible to seamlessly switch back and forth from the webpage in the DL to the visualization-based interface. The hypothesis was that by replacing the traditional webpage-based paradigm with a suite of visualizations that expose various dimensions of bibliographic meta data a researcher could gain new insights and find relationships that were not previously apparent. Furthermore, by making the experience spatial, the aim was to see if researchers were able to find what they were looking for more quickly and intuitively. The available bibliographic meta data in a DL is very interesting to explore and by looking at papers' references and citations, a researcher should be able to easily explore a field of study and get acquainted with what papers and authors are the most important in a given field.

2 Goals and Requirements

The main goal of this work was two-fold. First, to develop and test the effectiveness of an application, *PaperCube*, that allows a scholar to interact with a DL and explore bibliographic meta data using a defined set of visualizations. These visualizations had to provide different levels of visibility into the bibliographic relationships for papers and authors yet maintain *Focus+Context* [4].

- **Navigate Paper Citation Network Relationships**

A paper's citation network had to be navigable in both directions of the citation relationship, meaning that it is possible to see not only the papers that a paper has referenced, but also what papers have cited it. Also, users needed to be able to see author citation and collaboration relationships that can be implied from the paper's bibliographic meta data.

- **Provide a Suite of Visualizations**

PaperCube had to incorporate a suite of visualizations that provide for different perspectives and levels of detail of paper and author relationships. The visualizations needed to allow for the easy representation of not only a paper's immediate relationships, but also its extended citation network. Author relationships needed

to show interconnected author to author citation and collaboration relationships.

- **Switch Between Views Without Losing Focus**

PaperCube needed to be able to seamlessly switch between views that show the same paper or author and not lose focus on the item when the view is changed.

- **Allow User to Adjust the Amount of Data Shown**

The visualizations had to be easily customizable with a set of threshold parameters that constrain the amount of data rendered. Therefore, PaperCube needed to have UI slider controls that a researcher could adjust and, in turn, have the views update automatically in real time.

- **Resolution Independence**

All of the views needed to be resolution independent so that any portion of a visualization can be dynamically zoomed in to reveal more detail.

- **Use Existing Bibliographic Meta Data**

The purpose of this work was not to create a new set of bibliographic meta data. Therefore, the use of an existing DL with publicly available meta data that was easily searchable was needed.

- **Validation Through A User Study**

PaperCube needed to be validated through a user study that assessed the effectiveness of its design goals and methods. The study should assess the efficacy of the major design goal as well as the individual views and visualization methods.

The second goal was to push the limits of modern web browsers. By using web standards-based technologies, the goal is to explore the possibility of creating a dynamic, desktop-like experience that incorporates rich, interactive visualizations. The web is the best way to disseminate information, and it would be a great disservice not to use it.

3 Related Work

Previous research into the visualization of bibliographic meta data has evolved over the past twenty-five years as computers have become more powerful and capable. Several older tools serve as some of the inspiration of PaperCube, namely the *Document Lens* [14], *Butterfly* [11], and *BIVTECI* [12]. These tools all show similar bibliographic meta data, but use different visualization methods.

More recent tools such as *BiblioViz* [15] and *PaperLens* [9] stemming from InfoVis competitions have attempted to visualize paper and author relationships using newer rendering technology. *BiblioViz* uses both 2D and 3D visualizations to show paper publishing and author clustering. *PaperLens* used 2D visualizations to show citation and

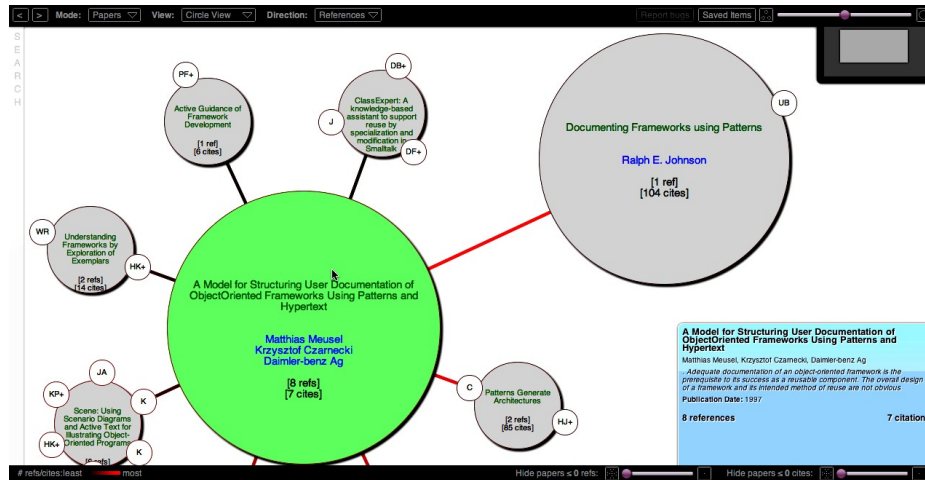


Figure 1. Screenshot of *Circle View* zoomed in.

collaboration connections throughout a community of researchers. Furthermore, *Circle View* [2], developed by Peter Bergström in 2004, was a novel visualization method to show a paper’s immediate citation network.

PaperCube expands upon previous DL visualization tools by incorporating the visualization of paper citation networks as well as the author collaboration relationships. Furthermore, by using a suite of visualizations methods including fisheye views [4] and tree maps [8] implemented to be web-based, the application is easily used by researchers.

4 PaperCube Description

PaperCube is a “cloud” or thick-client application that runs in a user’s web browser using SproutCore [7], a JavaScript-based UI framework. Rendering of visualizations is accomplished through the use of HTML, CSS, and Scalable Vector Graphics [3]. PaperCube allows a researcher to explore a version of the CiteSeer [5] data set and browse paper and author relationships through a set of highly interactive and dynamic, interconnected visualizations. Both of the paper and author modes have “detail” views that present the data in the visualization-based views in a table-based format that serves as a jumping off point to the other views.

4.1 Paper Visualizations

The main purpose is to explore the augmentation of DL search and browsing through the use of visualizations of paper citation network relationships. As a result, a set of tasks were envisioned and views were created to accomplish those tasks. The hypothesis is that creating a set of visualizations to accomplish those tasks will make Paper-

Cube very effective helping the users of DL services find what they are looking for faster and more intuitively.

The first task is the exploration of the paper’s immediate citation network in both direction of relationships, references and citations. This task would be useful once a paper has been found to find other papers that were immediately referenced by the authors. These papers are usually strongly related to the subject matter of the focused paper. Two visualizations were created, an improved version of *Circle View*, shown in fig. 1, and *Paper Graph*.

The second task is the exploration of the paper’s extended citation network viewed hierarchically. This task allows the researcher to explore deeper into in the citation network of a paper. The papers that are cited on the shallowest levels will be very targeted to the focused paper’s subject, but the more hops away from the focused paper, the less the papers may be directly related, yet still important to a researcher’s paper search. Papers that may seem unrelated at first can ultimately be very beneficial to a user’s research. To accomplish this, *Tree Map* view (fig. 2) was created.

The third task is looking at a focused paper’s citation network as it has developed over time, which may perhaps show temporal relationships that may otherwise be difficult to notice. Charting the evolution of a field from the most recent research back to a seminal work can be very interesting not only to find other papers in the field, but also to learn where it belongs in the overall tree of knowledge. As a result, the *Papers Per Year* view, shown in fig. 3, was created as a novel way to approach this problem.

4.2 Author Visualizations

As a secondary goal, this work wanted to explore how the author relationships could be interpreted through visualization as a way to allow researchers to find papers via

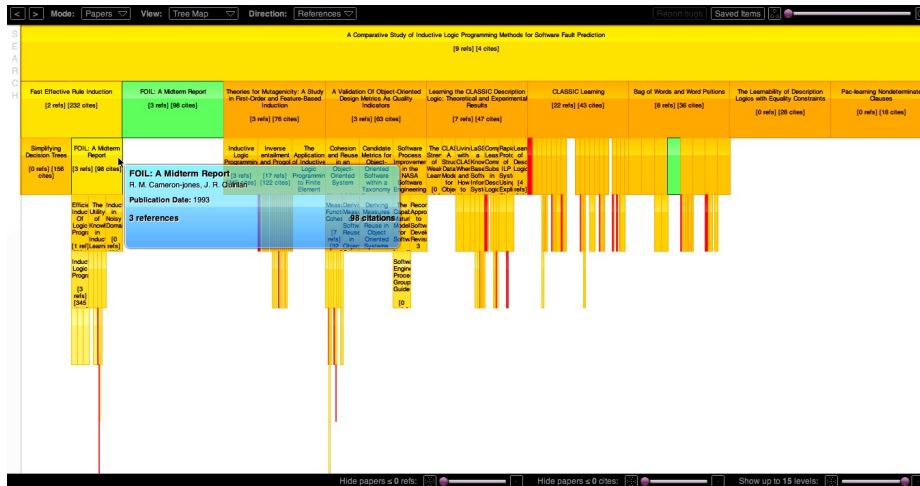


Figure 2. Screenshot of Tree Map view.

a different dimension on the available bibliographic meta data. Most existing digital libraries are solely focused on paper relationships and being able to derive and visualize author relationships from the existing meta data would give researchers a novel way to find new papers. The goal was to show author relationships that could be directly inferred from paper relationships that exist in digital libraries. A paper contains information about its authors and from this, the relationships between authors can be derived.

The first task was finding new authors from looking at an author’s collaborators. The hope was that showing collaboration relationships between authors could be a powerful feature and appreciated by researchers. The view, *Collaborators*, is shown in fig. 4. By finding a paper and looking at its authors, it should be very useful to look at the author’s collaborators and then find other important papers.

Second, it was thought that it would be interesting to know what authors have cited a given author and what authors that author had referenced across all his or her published papers. The hope was to create a view that gives a general overview of the citation space throughout the career of an author. The potential for this view was that it would be easy to see if an author influences many authors or vice versa. The view, *Author Cites*, is shown in fig. 5.

5 User Study

A user study was conducted asking past and present graduate students, researchers, and professors to participate. The survey’s goal was to get their thoughts on current web-based digital libraries, the general features and interface elements used in PaperCube, as well as rating the effectiveness of the various paper and author visualizations used. In order to provide the participants with some context, a brief video

[6] was created to give the participants some background of the goals of PaperCube as well as showing the available features and views. Full results can be found in [1].

The user study showed that PaperCube has a lot of potential and the participants’ impression of it overall was excellent. Eighty-five percent of participants found that PaperCube was useful when it comes to augmenting web-based digital libraries and 100% of the participants said that they would use it as part of a larger DL service.

In general, the survey validated top-level design choices of PaperCube. The decision to show both papers and authors views was well-liked by the participants and as expected, the paper views were found more useful than the author views—87% versus 13%. Also, the ability to adjust the display parameters of the views in real time was rated very highly. Furthermore, the zooming and overall feature of resolution independence was very well liked by the participants.

Rank	Papers	Pct	Authors	Pct
1	Circle View	32%	Author Detail	40%
2	Paper Details	29%	Collaborators	37%
3	Papers Per Year	24%	Author Cites	17%
4	Tree Map	9%	Papers	7%
5	Paper Graph	6%	-	-

Table 1. Most liked views in PaperCube

Shown in table 1, the most liked paper view was *Circle View*. It was surprising that *Paper Detail* view rated second highest followed by *Papers Per Year* view. The original hypothesis was that *Tree Map* view would have been rated highest or next to highest but in fact was second to last. The major reason cited by participants was that the view displayed too much information. *Paper Graph* view

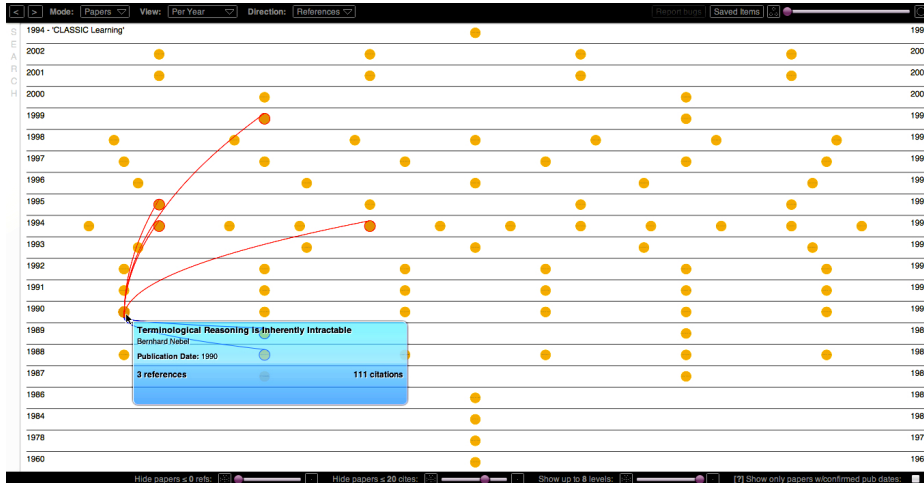


Figure 3. Screenshot of *Papers Per Year* view.

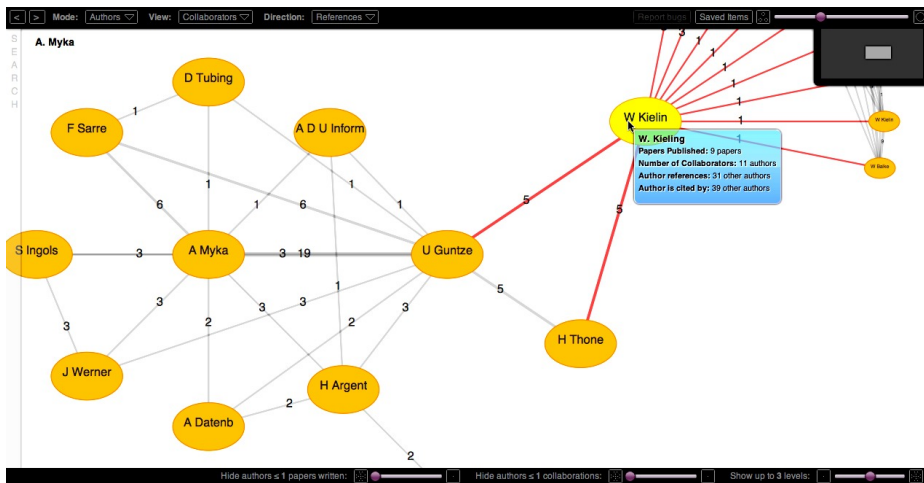


Figure 4. *Collaborators* view zoomed in showing the collaboration network of an author up to 3 levels.

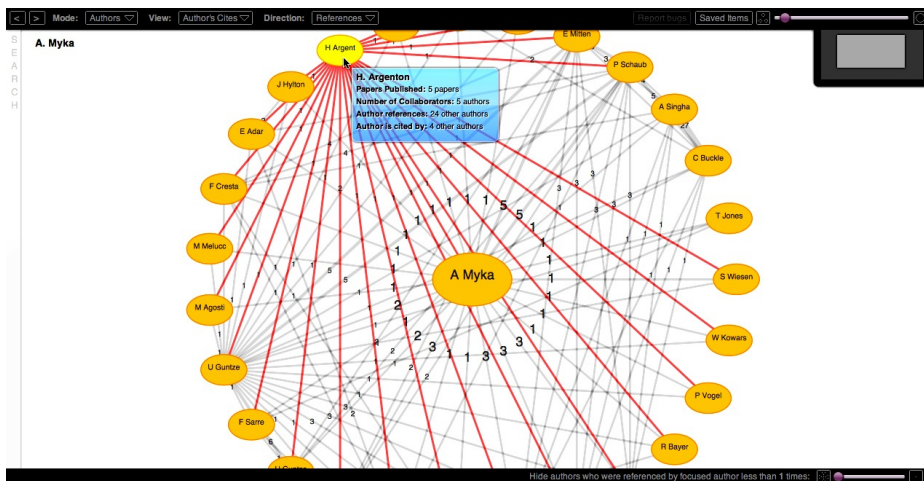


Figure 5. *Author Cites* view showing the authors that the focused author has referenced.

was rated the lowest because it was somewhat redundant to *Circle View*.

Also shown in table 1, it was expected that the most liked author views would be *Collaborators* view followed by *Author Detail* view. However, the results show that *Author Detail* view, which was picked by 40% of participants which slightly edged out *Collaborators* view. As expected, the two other views, *Author Cites* and *Papers*, were rated lower.

It was shown that PaperCube would fit very well with a DL service because most it helps reduce the “cognitive load” on the user by only showing relevant information. Also, showing author and paper relationships “visually” is a good way to increase the “discoverability” of new material. PaperCube is a “novel way” to view the relationships between papers in a way that is much more “understandable than the flat views” in most web-based digital libraries.

PaperCube would make “references searches much faster and complete” and as one participant said, it would be a “fantastic way to navigate and explore to find serendipitous connected papers/studies, and would make a particular library’s catalog that much more valuable.” Also, one participant especially liked the fact that the ability to filter the visible data was “much more extensive” than have other digital libraries allow.

The user study showed that PaperCube has a lot of potential and the participants’ impression of it overall was excellent. The participants rated PaperCube highly when it comes to augmenting web-based digital libraries and all of the participants said that they would use it as part of a larger DL service. Although not explicitly mentioned, the participants recognized that the goal for PaperCube was to be a potential companion to a full-featured DL service.

5.1 Discussion

The user study showed that PaperCube’s effectiveness stems from the views of the bibliographic meta data that are exposed to the user. Although showing only direct relationships between papers and authors, the types of visualizations used were in general thought to be effective. Users could focus on a paper or author at the low level or explore the networks of relationships up to fifteen levels deep.

The user survey showed that participants liked to see the details of a paper or author first, then branch out using the other views starting with the ones showing the most immediate relationships then going beyond. Therefore, *Circle View* and *Collaborators* view, visualizations that were more constrained in scope, were rated the highest by participants. Views such as *Papers Per Year*, *Tree Map*, and *Author Cites* were not rated as highly because participants thought that the potential data density was overwhelming at first.

Furthermore, the study showed that PaperCube made it possible to find references, citations, and authors that are

not directly connected with the focused paper or author through the ability of seeing many levels of relationships at once. Therefore, a researcher can stumble upon important papers and authors that may seem unrelated to the direct search at hand, yet indirectly be at the heart of the researcher’s search.

5.2 Opportunities For Improvement

Based on the survey results, several possibilities to improve PaperCube were found. First, improving the UI design to make it more intuitive would be very useful. Most of the improvements would be to make the color scheme a bit more pleasing and some of the UI elements stick out more. The aim was to make the UI fade away and allow for minimal distraction while the user is interacting with the various views. However, the subtleness of the interface made PaperCube have a higher learning curve to the novice user than expected. The participants noted that once the learning curve was overcome, PaperCube was easy to use.

Furthermore, participants thought that instead of having more views, PaperCube might benefit from having less views but with more features. By having more targeted views, PaperCube could be a lot more focused and easier to use. Not all of the views were as useful as they could have been. One prime example is the *Papers* view showing all the papers that an author has published. That view could be removed and improve PaperCube as a whole by making it more focused.

6 Future Work

PaperCube as validated in the survey is a self-contained application using a static data set. The participants noted that they would like to see PaperCube interact with a live data set. The interaction model with a live DL could manifest itself in two ways.

First, one could give PaperCube more search capabilities to search a DL directly and exist as a stand-alone service. This would have the benefit of being self-contained and only the interface to the DL would have to be developed. However, the rich capabilities of an existing web-based DL would have to be duplicated in PaperCube which might distract from its central goal, namely focus on the visualization of bibliographic meta data. Adding additional features to make it a full interface to a DL would be a distraction.

Second, one could make PaperCube even more lightweight and incorporate it into a web-based DL directly. The idea is that when a user views a paper in a DL a link would be present to show that paper in PaperCube. From PaperCube, the user would be able to navigate and view papers and authors like in the existing application.

The second interaction model with a DL might be the most desirable and effective. PaperCube was designed to augment existing digital libraries through the use of visualizations, not replace them.

7 Conclusion

This work set out to develop and test the effectiveness of an application that allows a scholar to interact with a DL and explore bibliographic meta data using a defined set of visualizations. This application, PaperCube, was designed to augment—not replace—existing DL services. PaperCube uses a set of visualizations to allow a scholar to be focused on a paper yet see where it belongs in the greater context of a publication space. Current digital libraries are limited by the webpage-based paradigm accepted as the norm. PaperCube set out to push the limits of web browsers and see if it was possible to break the mold and use a new paradigm to navigate bibliographic meta data. This new paradigm went beyond and created a dynamic, desktop-like experience that incorporated rich, interactive visualizations.

The user study validated that PaperCube’s set of views achieved the goal of augmenting digital libraries effectively. Participants unanimously said that PaperCube would fit well within the framework of an existing DL service and that if available, they would use it. Although the data set used focused on papers in the engineering and scientific fields, participants from other disciplines such as law found that with the right data set, PaperCube would greatly improve their daily workflow.

Participants found that PaperCube as a whole helped reduce the cognitive load of a researcher by making it easy to focus and show relevant information. All the views gave the user the control to adjust the amount of data shown at any time by a set of parameters. Especially in data dense views, this control made it easy to determine what papers or authors are important by altering significance factors. Therefore, instead of trying to automatically determine what is relevant, PaperCube trusts the user to determine relevance.

This work showed that augmenting the foraging for research material in DL discover through visualizations on the web is a possibility. PaperCube was shown to be to reduce the “cognitive load” put on a scholar and aid the “discoverability” of new research material. Furthermore, it was shown that participants thought that it was “visually exciting and intuitive” application and an “amazing example of the apps that we’ll be seeing on the web in a couple of years.”

References

- [1] P. Bergström. Augmenting digital libraries using web-based visualizations. Master’s thesis, Santa Clara University, Department of Computer Engineering, June 2009.
- [2] P. Bergström and E. J. Whitehead. Circleview: Scalable visualization and navigation of citation networks. In *Proceedings of the 2006 Symposium on Interactive Visual Information Collections and Activity (IVICA 2006)*, College Station, Texas, 2006.
- [3] J. Ferraiolo, F. Jun, and D. Jackson. Scalable vector graphics (SVG) 1.1 specification, W3C recommendation. Technical report, World Wide Web Consortium, 2003.
- [4] G. W. Furnas. Generalized fisheye views. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 17(4):16–23, 1986.
- [5] C. L. Giles, K. D. Bollacker, and S. Lawrence. Citeseer: An automatic citation indexing system. In *DL ’98: Proceedings of the Third ACM Conference on Digital Libraries*, pages 89–98, New York, NY, 1998. ACM.
- [6] Introduction to PaperCube. <http://vimeo.com/3323956>, 2009.
- [7] SproutCore: A JavaScript framework. <http://www.sproutcore.com>, 2009.
- [8] B. Johnson and B. Shneiderman. Treemaps: A spacefilling approach to the visualization of hierarchical information structures. In *VIS ’91: Proceedings of the 2nd Conference on Visualization ’91*, pages 284–291, Los Alamitos, CA, 1991. IEEE Computer Society Press.
- [9] B. Lee, M. Czerwinski, G. Robertson, and B. B. Bederson. Understanding research trends in conferences using paperlens. In *CHI ’05: CHI ’05 Extended Abstracts on Human Factors in Computing Systems*, pages 1969–1972, New York, NY, 2005. ACM.
- [10] C. Lynch. Where do we go from here? The next decade for digital libraries. *D-Lib Magazine*, 11(78), 2005.
- [11] J. D. Mackinlay, R. Rao, and S. K. Card. An organic user interface for searching citation links. In *CHI ’95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 67–73, New York, NY, 1995. ACM Press/Addison-Wesley Publishing Co.
- [12] D. Modjeska, V. Tzerpos, P. Faloutsos, and M. Faloutsos. BIVTECI: A bibliographic visualization tool. In *CASCON ’96: Proceedings of the 1996 Conference of the Centre for Advanced Studies on Collaborative Research*, page 28. IBM Press, 1996.
- [13] National Science Board (U.S.) and United States. *Science and Engineering Indicators 2008*, volume 1. Washington, U.S. Govt. Print. Off., 2008.
- [14] G. G. Robertson and J. D. Mackinlay. The document lens. In *UIST ’93: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, pages 101–108, New York, NY, 1993. ACM.
- [15] Z. Shen, M. Ogawa, S. T. Teoh, and K.-L. Ma. Biblioviz: A system for visualizing bibliography information. In *APVis ’06: Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation*, pages 93–102, Darlinghurst, Australia, 2006. Australian Computer Society, Inc.