

# Augmenting Digital Libraries with Web-Based Visualizations

Peter Bergström, Darren C. Atkinson  
Santa Clara University, USA



Journal of Digital  
Information Management

**ABSTRACT:** *Digital libraries in their current form are bounded by their inefficient webpage-based user interface paradigm, and even the most knowledgeable researchers can get lost in the large amount of published material available. A paper and its immediate references are displayed on a single webpage. Unfortunately, it is not readily apparent where the paper belongs in the greater context of a research field. Our goal was to develop, test, and iteratively improve the effectiveness of a web-based application, PaperCube, designed to augment and enhance a researcher's interaction with digital libraries through the use of interactive visualizations of bibliographic meta data. The visualizations set out to show different levels of detail and aspects of a paper's citation network. A primary concern was to ensure that while switching visualizations, the viewed paper or author remained focused at all times. PaperCube was validated through a user study that showed that it helped users who are trying to explore a field of research. In particular, PaperCube was shown to reduce the "cognitive load" put on a researcher and aid the "discoverability" of new research material.*

## Categories and Subject Descriptors

**H.3.5 [Online Information Services]:** Web-based services; **H.3.7 [Digital Libraries]:** System issues

**General Terms:** Visualization; Digital Libraries; User interfaces; Metadata

**Keywords:** Digital libraries, Navigation, Web-based visualization, Graphical user interfaces, PaperCube

**Received:** 12 March 2010; **Revised** 19 April 2010; **Accepted** 11 May 2010

## 1. Introduction

As a researcher sets forth to develop new and novel concepts, ideas, and inventions, he or she more often than not expands upon previous published research. Therefore, a researcher needs to search for related publications in his or her field. Once the researcher has found a set of articles, papers, or books, the next step is to explore other works through their references. Even if the person is an established researcher in a given field, finding valuable and pertinent material can be a difficult and time-prohibitive process.

The publication rate of research papers has been accelerating in recent years. As a result, scholars have to sift through too much published material of varying quality to find research relevant to their topic of interest. The National Science Foundation (National Science Board, 2008) noted that the global publication rate of science and engineering papers has grown at a steady yearly rate of 2.3 percent between 1995 and 2005. Furthermore, between the years 1988 and 2005, the amount of interinstitutional collaboration between authors grew from 40 percent to 61 percent. As a consequence, it is pivotal to develop new and improved methods to help researchers through the use of technology, particularly on the web.

## 1.1 Digital Libraries

Digital libraries are not a new concept that were created with the help of the web. Starting in the 1960s and 1970s, standalone systems and systems that used the early Internet were developed (Lynch, 2005). Unfortunately, these systems were very specialized and, due to limitations of the technology of the time, inaccessible to most researchers. However, with the popularization of the web in the mid-1990s, web-based digital libraries were developed.

Although the web has made it easy to search and find publications, a new challenge has arisen: how to effectively manage and alleviate information overload. It is easy for a researcher to get lost in the clutter of irrelevant publications and never find any papers of actual value to his or her interests. The search features offered by contemporary digital libraries make it simple to narrow down the domain of papers, but the webpage-based user interfaces used today are inadequate. The current user interface paradigm of the web constrains the user and results in a poor user experience: searching, viewing, and foraging for papers is through the navigation from one webpage to another. As a result, it is difficult to see where the user is coming from and where he or she may go next. A single paper in a digital library is represented as a webpage that contains links to other papers that reference it. It is not simple for a user to see where the paper belongs in the wider context of the digital library and, without navigating away from the paper that is being viewed, it is not possible to view second order references, and so on, or in other words, the greater citation network. In order to view a paper's reference, the researcher needs to navigate away from the current page or open a new browser. Unfortunately, both of these solutions result in poor usability because the researcher now has lost the context of the original paper he or she was viewing.

As a result, it is important to create systems that allow for the navigation of bibliographic meta data where a paper can still be focused on screen and still facilitate the easy and seamless retrieval of the surrounding citation network to provide context of where the focused paper belongs. Such an application does not necessarily need to replace contemporary web-based digital libraries, but rather just augment them to provide additional information and context that a webpage cannot provide by itself. However, it is important to remain web-based so that it is possible to seamlessly switch back and forth from the webpage in a digital library to the visualization-based user interface. The hypothesis in this work is that by replacing a traditional webpage-based representation of a paper within a digital library with a suite of visualizations that expose several different perspectives of the same bibliographic meta data, a researcher could gain new insight and find previously non-apparent relationships. By making the user experience focused on representing bibliographic relationships spatially, the goal

was to see if researchers were able to more quickly and effectively find what they were looking for. The meta data available in a digital library is very interesting to explore and by looking at the relationships between papers and authors, a user should be able to explore a research domain and find who and what are the most important in a field.

## 2. Goals And Requirements

This work had two main goals. The first goal was to develop and experimentally validate the effectiveness of a web-based application, *PaperCube*, that allows a researcher to interact with a digital library and explore the bibliographic meta data that it contains using a suite of visualizations. As stated earlier, these visualizations are meant to show different levels of detail of paper and author relationships yet maintain proper *Focus+Context* (Furnas, 1986).

### • Navigate Paper Citation Network Relationships

The citation network of a paper had to be navigable both forward and backward, meaning that it is possible to view not only the papers that a focused paper has listed as its references, but also what other papers list that focused paper as a reference. Furthermore, users of the application should also be able to view author citation and collaboration relationships that could be extracted from the available bibliographic meta data.

### • Provide a Suite of Different Visualizations

The application had to contain a suite of visualizations that provided for various perspectives and levels of detail of paper and author relationships. The visualizations needed to represent not only a paper's immediate reference relationships but also its extended citation network. Author relationships needed to show author-paper citation relationships as well as author-collaboration relationships.

### • Switching Between Visualizations Without Losing Focus

Another important goal was to allow the user to switch between different visualizations seamlessly to see different dimensions on the same paper or author without losing focus. When switching between views, the viewed item should not change and no additional action should be taken by the user to view an item in a different view other than switching the view option.

### • Allow User to Adjust Visualizations in Real Time

Since there is the potential to display a very large amount of data on the screen, the user should be able to narrow down the visible data to suit his or her needs by using slider controls in the views to manipulate a set of independent display parameters. The visualization should allow for the depth of the visualization to be altered in real time as well as change the visualization based on other threshold constraints on the data. For example, the user should be able to hide papers by the number of references or citations they have.

### • Use Existing Bibliographic Meta Data

The purpose of this research is not to create a digital library. Therefore, it is important to find a suitable digital library that already exists, and mine its bibliographic meta data so it can be used in the tool. Therefore, the plan was to use the CiteSeer data set.

### • Validation Through a User Study

*PaperCube* needed to be validated through a user study. This study needed to assess the effectiveness of its design and

implementation as well as evaluate the efficacy of 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 was to explore the possibility of creating a dynamic, desktop-like experience that incorporates rich, interactive visualizations.

However, since the goal of this project was to explore new ways to represent the data, a more desktop-like experience was needed, yet still remain web-based. Also, even on the desktop, it can be difficult to create an interactive application with rich visualizations and make it perform well. In pursuit of creating something that is forward-looking, it was decided upfront that the application would use open standards. Therefore, the application is written using JavaScript, Scalable Vector Graphics, Canvas Tag, CSS, and HTML. However, this did present a unique challenge because of cross-browser issues. To alleviate this, open source browsers were supported first and foremost. Thus, Firefox, WebKit, Safari, and Google Chrome are fully supported but Internet Explorer is not.

## 3. Related Work

The visualization of bibliographic meta data has evolved over the past three decades as computers became increasingly powerful. Older tools such as *Document Lens* (Robertson & Mackinlay, 1993), *Buttery* (Mackinlay et al., 1995), and *BIVTECI* (Modjeska et al., 1996) served as inspiration for *PaperCube*. These tools all show similar bibliographic meta data but through differing visualization methods.

Within the context of this work, *Circle View* and, to a limited extent, the NodeGraph-based views are a kind of fisheye view like that of *Document Lens*, but rather than using a rectangular viewing context, they use circles to accomplish something similar. However, one significant difference is that the *Document Lens* does not show relationships, but rather allows the user to browse a large amount of unstructured information.

At the core of *Butterfly's* user interface, there is a "butterfly," a three dimensional representation of a paper's bibliographic meta data. One "wing" is used to list an article's references and the other "wing" is used to list the article's citers. *Butterfly* uses the idea that an article with many citers is probably a good article to highlight to the user, a concept that is used in *PaperCube*.

*BIVTECI* addresses the need for an interactive bibliographic visualization tool that allows users to access the complete bibliographic data for an article, filter by title, author, and keyword, ordered citation links, information views at several levels of detail, multiple synchronized views, and the visualization of large data sets. *PaperCube* shares this methodology of showing different levels of detail across many synchronized views which allows the user to see the greater context where a paper or author belongs without losing focus.

There are also tools that were developed more recently: *BiblioViz* (Shen et al., 2006) and *PaperLens* (Lee et al., 2005) came to fruition through InfoVis competitions and attempts to visualize paper and author relationships using recent rendering technology and techniques. *BiblioVis* uses both two and three-dimensional visualizations to display paper publishing and authorship clustering. *PaperLens*, on the other hand, uses two-dimensional visualizations to show citation and collaboration relationships throughout a community of scholars.

*PaperCube* is itself a direct descendant of *Circle View* (Bergström & Whitehead, 2006) and expands upon previous digital library visualization tools by incorporating the visualization of citation

networks and author collaboration relationships. By using a set of views inspired by visualization and user interface design methods including fan menus (Hopkins, 1991), fisheye views (Furnas, 1986), hyperbolic trees (Lamping et al., 1995), and tree maps (Johnson & Shneiderman, 1991) implemented using native web-based technologies, PaperCube can easily be used by researchers.

#### 4. Papercube Description

PaperCube, shown in fig.1, is an application that allows a researcher to explore a version of the CiteSeer (Giles et al., 1998) digital library and browse paper and author relationships using a set of highly interactive and dynamic, interconnected visualizations. Both of the paper and author visualization modes have “detail” views, which present the available meta data presented in the visualizationbased view in a table-based presentation format that serves as a launching pad to the other views.



Figure 1. The PaperCube user interface on initial load

Although PaperCube is web-based, the aim was to create a highly interactive application that allows researchers to interact with it in a similar manner to a desktop application. The idea is to make the fact that PaperCube resides within the confines of a web browser almost impossible to tell. This “cloud”—or more technically—thick-client web application architecture is possible through the use of SproutCore (Sprout Systems Inc., 2009), a JavaScript framework that is specifically tailored to deliver desktop-class user interfaces on the web. The visualizations are rendered through the use of HTML, CSS, Canvas Tag (Hickson, 2009), and Scalable Vector Graphics (Ferraiolo et al., 2003). A more indepth technical discussion regarding the implementation of PaperCube was covered by the authors in a paper for SVG Open 2009 (Bergström & Atkinson, 2009).

Just as a desktop application, PaperCube loads once per session and any subsequent actions are handled without reloading. This allows for rich, immersive user interaction that is fast and seamless. Communication with the server takes place asynchronously in the background and allows a researcher to interact with PaperCube without interruption.

#### 5. Searchable Data Set

The need for a large set of bibliographic meta data was obvious for an application such as PaperCube. Initially, a basic programmatically generated data set was used to begin development. However, a synthetic data set does not necessarily model the makeup the bibliographic meta data contained in an actual digital library. In order to test PaperCube's effectiveness

at augmenting current webpage-based digital libraries, it was necessary to find a real data set. The most logical choice was CiteSeer<sup>x</sup> because it provides an OAI (Lagoze et al., 2002) interface for harvesting its meta data. However, the harvested CiteSeer<sup>x</sup> bibliographic meta data did not work correctly because the reference relationship links between papers were incorrect or incomplete. During *Circle View's* development in 2004, the original CiteSeer meta data was explored but dropped in favor of a limited set of bibliographic meta data from the ACM Digital Library.

However, for this work, it was necessary to test a larger set of bibliographic meta data and therefore, the original CiteSeer data from 2004 was used. It was augmented for correctness by using additional information from CiteSeer<sup>x</sup>. Also, in order to allow for the exploration of authors, the bibliographic meta data set was mined to get additional author relationships such the papers published, collaborators, the authors that an author has referenced and the authors that have cited the author. Even though the harvested CiteSeer<sup>x</sup> meta data did not yield usable reference links, it did provide correct information regarding publication year for half of the data set.

Currently, PaperCube allows a researcher to search and explore the bibliographic meta data for 716,772 papers, 411,034 authors, and 1,764,929 reference links. Although outdated, the set of bibliographic meta data used should be sufficient for the purpose of validating PaperCube.

#### 5.1 Navigation “Fan” Menu

One challenge that was encountered during development was that in information dense visualizations it is paramount to reduce the amount of additional clutter that may distract from the visualizations themselves. In each view, when a node is moused over, associated meta data is displayed in the form of a hovering information pane that follows the mouse cursor. In order to allow for actions to be applied to nodes in a visualization, these actions would have to be implemented by clicking on the node.

As a result, it was decided to create a pop-up menu that provides a set of actions that can be applied to an item. Several UI design concepts were explored. First, a drop down menu that listed choices in a list was considered but it was thought to be relatively space inefficient. Also, it was thought that such a menu would be slow since the mouse has to move far to select a menu choice. Instead, the most interesting solution seemed to be to use a fan menu, also known as a marking or radial menu. Fan menus are very efficient because the mouse cursor

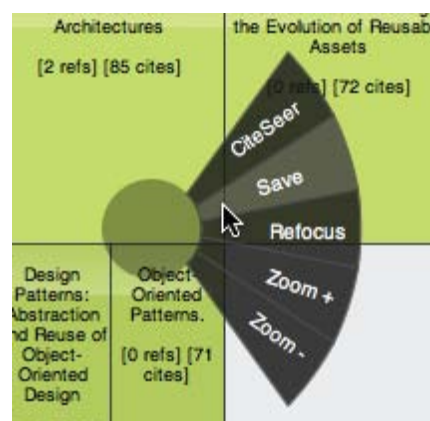


Figure 2. The fan menu control activated on a paper in the Tree Map view

does not have to move far to select an option in the menu. The menu is positioned so that it surrounds the mouse cursor. Around the mouse cursor, a set of actions are displayed that can be quickly selected.

In PaperCube, when a node is clicked with the left mouse button, the fan menu is displayed. The options in the views depend on the type of data being displayed. As shown in fig. 2, the available actions for a paper allow for the zooming in or out of the visualization, saving the clicked paper, refocusing the view on the paper, and view the paper at the source digital library, CiteSeer.

## 5.2 Slider Controls

Instead of having the views attempt to automatically determine the optimal amount of data to be shown for a paper or author in a view, PaperCube defers to the user to determine the rules that determine how much is rendered on the screen. These display parameters are threshold parameters that are used to constrain the amount of items that are visible on the screen. The hope was that the researcher using PaperCube knows best when it comes to determining what bibliographic meta data is the most important to show on the screen.

At the bottom of the browser window, with the exception of *Paper Detail* and *Author Detail*, views have a set of one or more slider controls. These controls allow the researcher to easily adjust the threshold parameters to hide or show more bibliographic meta data in order to tailor the view to show what is most useful to the individual. Some of the common thresholds that can be adjusted via the sliders include: the retrieval depth of a paper's citation network or an author's relationship network, the visibility of papers based on its number of references or citations and the visibility of authors based on the number of papers written or the number of collaborations with other authors. In addition to controlling the visual representation of the viewed bibliographic meta data, the sliders also determine the amount of data retrieved from the server, and the more strict your display parameters are, the less needs to be loaded from the server. If more information is revealed by changing a display parameter, only the bibliographic meta data that has not previously been loaded into the SproutCore data store will be retrieved from the server.



Figure 3. The slider controls for **Tree Map** view showing three different adjustable display parameters

## 5.3 Zoom Widget

Most of the views in PaperCube are designed with resolution independence in mind. Resolution independence was a top goal when creating the tool. Since the visualizations can be complex, zooming in to get more information can be very useful. In fig. 4, the zooming widget is located at the top right corner of the screen. It consists of two buttons and a slider control. The buttons either zoom in and or out by one step of the slider, and the slider can be used to dynamically zoom in and out by dragging the handle with the mouse. It is also possible to zoom in and out using the mouse scroll wheel or through the fan menu.

When zoomed in, a preview UI slides down that shows the full area of the visualization and the currently visible area. With the mouse, one can zoom and pan the whole visualization from the preview pane. When panning using the preview pane, the visualization does not update until the user lets go of the mouse button. It is possible to pan the visualization from

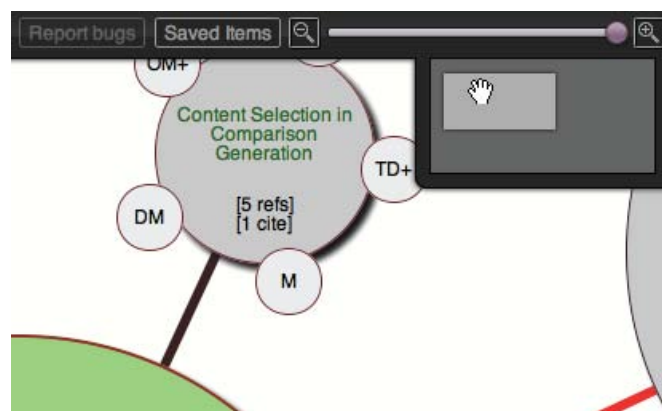


Figure 4. The zoom preview window is visible when zoomed in and can be used to pan around the visualization

the main window by grabbing the available white space and dragging the visualization. This method is not as fast in terms of rendering as using the preview window since it updates the visualization in real time but it is more precise and gives immediate feedback.

## 5.4 Meta Data View

In the majority of views, screen real estate is at a premium. However, the goal is to show as much bibliographic meta data about an item as possible. In the visualizations themselves, only limited information can be displayed, but a lot of other important meta data is present that should not be missed about the items. Therefore, when the mouse cursor is moved above a node, a box containing additional meta data is revealed that follows the mouse cursor. As shown in fig. 5, a paper item's meta data consists of the title, authors, publication year, and the number of references and citations. An author's bibliographic meta data consists of the author's name and other high-level statistics such as the number of papers published, the number of collaborators, the number of times the author references others, and the number of times that other authors reference the author.

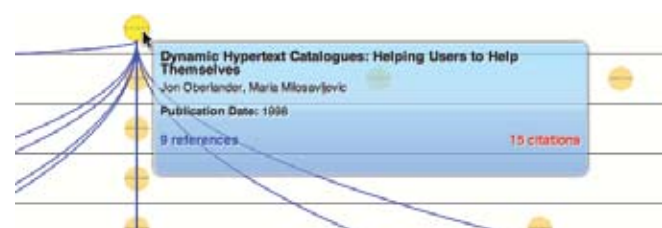


Figure 5. Meta data for a paper in **Papers Per Year** view

## 5.5 Search Pane

While the actual search for papers and authors is not the main focus of PaperCube, some limited searching functionality was required so that the tool is usable in its experimental form. The search pane is hidden on the left edge of the screen until the user mouses over the area.

The searching functionality allows bibliographic meta data about papers or authors to be retrieved from the database based on a paper's title, abstract, publication date, subject or an author's name. Most searchable fields allow the use of MySQL "full-text" query syntax (MySQL, 2009) that allows for powerful and precise searches. However, as previously mentioned, the functionality is limited and should not be compared to a fully developed digital library service.



## 5.6 Paper Visualizations

The main purpose of this work is to explore the augmentation of digital library 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 was that creating a set of visualizations to accomplish these tasks will make PaperCube very effective in helping the users of digital library services find what they are looking for faster and more intuitively.

**Showing a Paper's Immediate Citation Network.** The first task is the exploration of a 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, a refined version of *Circle View*, shown in fig. 6, and *Paper Graph*.

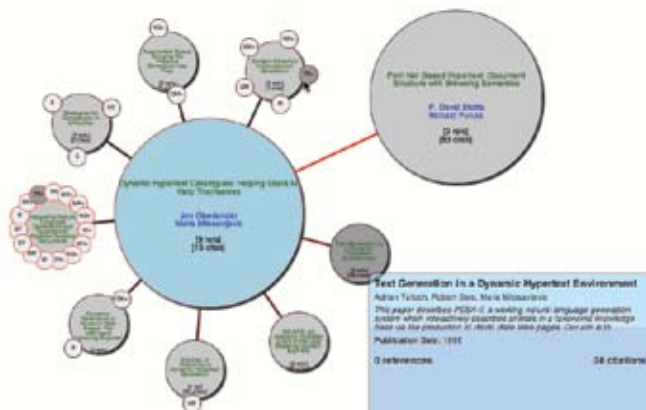


Figure 6. A screenshot of *Circle View*

**Circle View.** *Circle View* (Bergström & Whitehead, 2006) was originally created at University of California at Santa Cruz in 2004 as part of Peter Bergström's undergraduate thesis research under the supervision of professor E. James Whitehead.

The goal of this visualization method is to create a way to easily show a paper, its references and its references' references. The visualization used circles around circles and showed papers that occurred more than once in the graph as distinct nodes in order to eliminate the need to show edges crossing all over the screen. The visualization is arranged in a fisheye configuration with the focused paper showing a lot of meta data including its title, authors, number of references and citations. The subsequent levels of related papers are smaller and show less information. In order to provide a fast way to assess significance, color coding is used to signify the number of references and citations for the papers relative to the rest in the visualization. The color around the circles' border signifies the paper's number of references or citations and the lines that go from the focused paper to its first-level references signify the paper's number of citations.

When interacting with the view and mousing over a paper circle, bibliographic meta data pertaining to the paper is displayed and any additional occurrences of that reference are highlighted. Since the visualization allows for the same paper node to be rendered more than once, it is easy to count multiple occurrences on the screen. *Circle View*'s visualization method is very rigid and only allows for two levels of references or citations to be shown at any given time. Other views in PaperCube are much

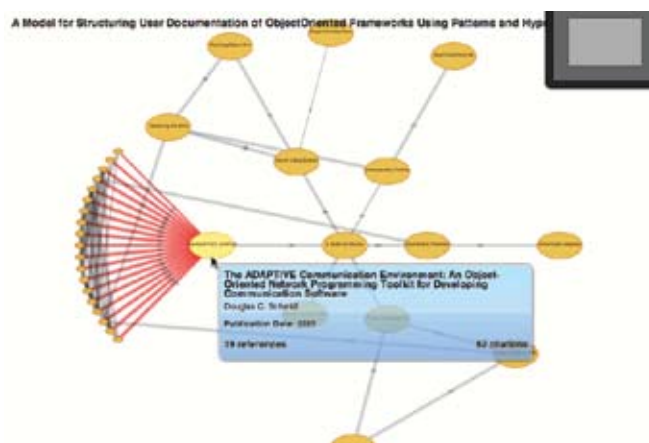


Figure 7. A paper's citation network using *Paper Graph* view

more flexible and allow for up to fifteen levels to be displayed.

**Paper Graph.** The motivation for creating *Paper Graph* view, shown in fig. 7, was to create a view of a paper's immediate citation network that offers the flexibility that *Circle View* explicitly eliminates by design. Each paper is represented as a node and any reference or citation relationships to other papers are shown as edges. The widths of the edges is variable depending on its implied relevance. In this case, the relevance is determined by the number of citations, when viewing its references, or the number of references, when viewing its citations.

As with the other views, it is possible to dynamically adjust the thresholds that determine the parameters of the visualization by hiding nodes that fall below the specified reference or citation thresholds. Also, in contrast to *Circle View*, the number of levels of the citation network is customizable.

**Visualizing An Extended Citation Network.** The second task is the exploration of a paper's extended citation network viewed hierarchically. This task allows the researcher to explore deeper into 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. As a result of this task, *Tree Map* view (fig. 8) was created.

When exploring ways to show a large number of levels in a paper's citation network, the best visualization method seemed to be a tree map. Tree maps are very powerful when it comes to showing deep, intricate hierarchical relationships. Therefore, the idea of using a tree map seemed logical. The hope for this view was to show as much bibliographic meta data as possible and let the user adjust thresholds that determine the display

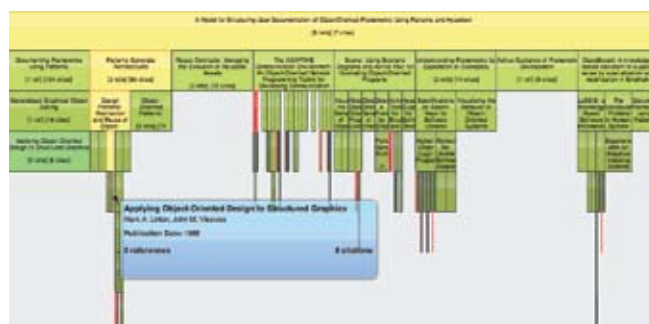


Figure 8. A screenshot of *Tree Map* view

parameters using slider controls to narrow down the potentially large data set to a very focused set of relevant information. The results from the user study pointed out that participants were overwhelmed by the large amount of information rendered.

**Visualizing Citation Networks Chronologically.** 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 useful not only to find other papers in the field, but also to see where it belongs in the overall tree of knowledge.



Figure 9. A screenshot of **Papers Per Year** view

The *Papers Per Year* view was created to accomplish this task using a novel approach, shown in fig. 9.

When viewing a paper's references, the focused paper is at the top of the browser window and when viewing a paper's citations, the focused paper is at the bottom. The view allows for the exploration of a paper's citation network and all the papers are represented as nodes organized chronologically by publication year.

## 5.7 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 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 bibliographic 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. Therefore, the goal was to visualize simple relationships and not use any calculations to determine what authors are more significant than other authors other than by looking at the number of relationships between authors and the papers that they have written.

**Finding Collaborators.** The first task identified is searching for new authors by looking at an author's collaborators. The hypothesis is that showing collaboration relationships between authors would be used by many researchers and creating such a visualization would be appreciated. The view, *Collaborators*, is shown in fig. 10. By finding an interesting paper and looking at its authors, one can look at the author's collaborators and then find other important papers.

**Finding Author Citations and References.** The second task that was identified is finding 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 cre-



Figure 10. A screenshot of *Collaborators* view zoomed in showing the collaboration network of an author up to more than 3 levels

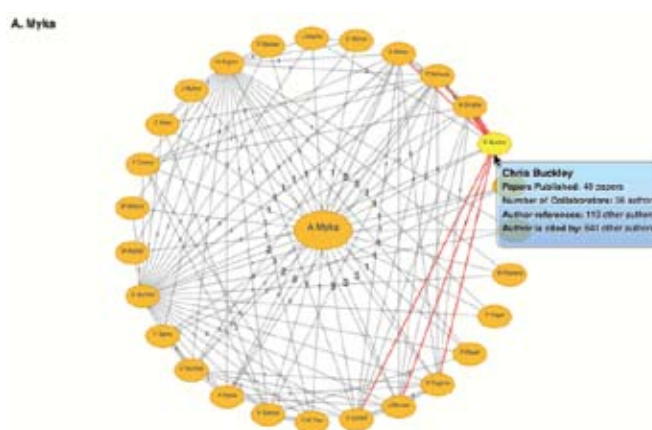


Figure 11. The **Author Cites** view showing the authors that the focused author has referenced

ate a view that gives a general overview of the citation space throughout the publishing 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. 11.

## 6. Papercube Architectural Overview

PaperCube consist of three distinct components. First, a MySQL database that stores the bibliographic meta data for papers and authors. Second, a PHP-based database interface listens to requests for paper or author bibliographic meta data from the client and returns the result from the database in JSON format. Third, the SproutCore-based application that runs in a web browser. The majority of the business logic and complexity lies in the web browser leaving the two layers below lightweight. Figure 12 shows a high-level architectural diagram showing the relationship between the three components.

## 7. User Study

A user study was conducted with past and present graduate students, researchers, and professors. The 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 visualizations.

In order to provide the participants with some context, a brief video (PaperCube, 2009a) was created to give the participants some background of the goals of PaperCube as well as showing

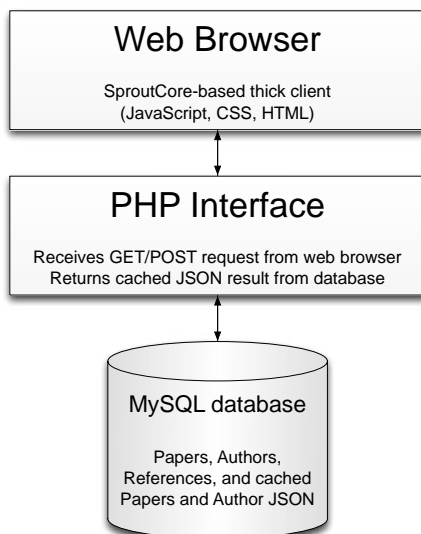


Figure 12. The high-level architecture showing the three distinct components of **PaperCube**

the available features and views. Full results and an expanded discussion can be found in Bergström (2009).

### 7.1 Study Goals

The goal for the study was not only to evaluate PaperCube's effectiveness at augmenting digital library search and navigation, but also to evaluate the specific user interface elements that were used throughout the application. Therefore, a set of HCI-centric questions were used to evaluate the various elements of PaperCube.

#### • PaperCube's Potential

Was PaperCube successful overall at augmenting the use of a web-based digital library and would this be useful to include in future digital library services? These questions were generally not in multiple choice format, but rather allowed the participant to respond freely.

#### • View and Visualization Effectiveness

The views and their visualizations are at the center of this work. The desire was to find out which views were better than others in the eyes of the participants. The visualizations put forth in PaperCube may or may not be as effective as hypothesized.

#### • Were Paper or Author Views Preferred?

Since PaperCube shows both papers and authors, which was preferred by the participants? What were the most effective views for each of the viewing modes?

#### • Slider Controls to Adjust Display Parameters

Was the ability of being able to adjust the display parameters of the visualizations useful to users?

#### • Zooming and Resolution Independence

Was the ability to zoom in on the visualization useful? The idea was that this would be a very nice feature especially when showing a lot of data on the screen. Was the preview widget at the top right of the window useful? How about dragging and panning?

### 7.2 Participants

The participant pool was made up of past and present graduate students, researchers with undergraduate degrees, and

professors. Forty-eight percent of participants were working professionals with a graduate degree. Their backgrounds were quite diverse, ranging from law school graduates, mechanical engineering master's students up to PhDs in computer science, computer engineering, and astronomy. The second largest pool of participants were graduate students pursuing degrees in various fields including computer science, computer engineering, law, geology, and library science.

### 7.3 Study Results

Eighty-five percent of the responses collected during the user study lead to the conclusion that PaperCube goes a long way to augment web-based digital libraries in a manner that is very useful to researchers. Furthermore, 100 percent of the participants said that they would use PaperCube as part of a larger digital library service offering. Overall, the study validated the top-level design choices that were made in PaperCube. The decision to pursue the visualization of both papers and authors was appreciated by the participants.

**Validating PaperCube's General User Interface.** One goal of the study was to validate if the UI features in PaperCube were effective. Therefore, a set of HCI-centric questions were asked so that participants could rate the effectiveness of some of the major functionality of PaperCube. These rating questions were numerically ranked from 1 to 5 where 5 was the most favorable rating.

Most participants rated PaperCube favorably with an average rating of 3.8 out of 5. Participants generally liked PaperCube and thought that it has "great potential." Especially, participants commented on its dynamic nature and one participant said that the "interface is surprising rich interactive and responsive" and another said that he is "impressed by the general responsiveness and the speed of data retrieval." Furthermore, several participants liked that PaperCube is very "fluid," "impressive," and "beautiful." Due to PaperCube's somewhat non-standard user interface, some participants had issues with some of the UI design decisions. Some participants found the UI "confusing" at first and that "too much functionality is hidden or not readily obvious." However, participants said that PaperCube does have a "learning curve," but "easy" once a user gets used to it.

One of the major meta-features of PaperCube was resolution independence. An overwhelming majority of the participants liked this feature, rating it an average of 4.4 out of 5. Participants thought that this was a very strong feature, calling it "excellent," "awesome," "the way it should be," and "fantastic!" One participant said that it was impressive but not sure if it has been fully "exploited" in PaperCube. The resolution independence used in PaperCube has a lot of potential. Furthermore, the use of "vector graphics" made it easy to scale fonts because normally it is really "messy" and this implementation keeps it "clean."

Another top goal of PaperCube was the ability to change the display parameters in order to change the amount of information visible on the screen. In general, the parameters that could be adjusted included hiding papers or authors based on a set of thresholds or adjusting the number of levels to be retrieved from the server. The aim of this feature was to allow the user to quickly narrow down what is being shown to something that may be relevant to that particular user. Just as with the zooming and resolution independence feature, the ability to adjust the settings of a visualization with the slider controls was overwhelmingly rated highly. The average rating for this feature was 4.2 out of 5 with 53 percent rating it a 5 out of 5. The majority of participants found this feature very useful. One

participant noted that this feature “lets the user narrow down results quickly” and another said that is a “great way to play around with the options.”

**Paper and Author Visualizations.** As hypothesized, the paper visualizations were found more valuable than the visualizations showing authors—87 percent versus 13 percent. The ability to dynamically adjust the visualization display parameters in real time was rated very high. Furthermore, zooming and the overall application-wide feature of resolution independence was highly valued by participants.

As shown in table 1, the highest ranked paper view was *Circle View*. This was not unexpected because of the prior positive response that this visualization had received at the 2006 Symposium on Interactive Visual Information Collections and Activity (Bergström & Whitehead, 2006). However, it was surprising that *Paper Detail* view rated second highest followed by *Papers Per Year* view. The expectation was that *Tree Map* view would have been rated much higher, but it was in fact rated second to last by participants. The majority of participants noted that *Tree Map* view displayed too much information. *Paper Graph* view was rated the lowest because it was seen as somewhat redundant to *Circle View*.

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. PaperCube paper and author views ranked by rating

As for author views, the highest rated view was *Author Detail* view. This view was picked by 40 percent of participants as their favorite. The *Collaborators* view, which was actually expected to be rated the highest, came in second with 37 percent. Not as surprising, the other two author views, *Author Cites* and *Papers*, were rated lower.

Participants noted that PaperCube would fit well within a larger digital library service because it helps reduce the “cognitive load” placed upon the user by making it easier to find relevant information. Also, showing author and paper relationships visually is a great way to increase the “discoverability” of material not previously known to a user. PaperCube is also a “novel way” to view the relationships between papers in a manner that is much more “understandable than the at views” commonly used in web-based digital libraries.

PaperCube enables “reference searches [to be] much faster and complete.” As one participant noted, PaperCube is 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 found the feature of being able to filter the visible data “much more extensive” than other digital libraries allow.

The user study showed that PaperCube is a useful tool and the participants’ overall impression of it was excellent. The participants appraised 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 greater, more broad digital library service. Although not mentioned to the participants, several remarked that PaperCube would fit in well as a companion to a full-featured digital library service.

## 7.4 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.

## 7.5 Possible Refinements

As a direct result of the user study, several avenues of improvement to the usability of PaperCube were found. First, improving the user interface design could make PaperCube more intuitive to use. The majority of the improvements would be to make the color scheme of the various views more pleasing and distinct. Participants said that it was difficult to distinguish several of the views from each other because they had the same color scheme.

An original idea used in the design of PaperCube was to make the navigation user interface elements 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 for the novice user than expected. The participants noted that once the learning curve was overcome, PaperCube was easy to use.

One specific example was the behavior of the search pane. By default, it was hidden on the left side of the screen. In order to lower the learning curve, having the search pane open when the application loads would be a simple enhancement and would result in a better user experience. Many participants reported that it was difficult to figure out where to start once PaperCube loaded.

Furthermore, participants commented that instead of having more views, PaperCube might benefit from having a smaller number views with more features. By having more versatile views, PaperCube could be 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 the user experience of PaperCube as a whole by not unnecessarily distracting users from more important views.

## 8. Future Work

PaperCube as validated in the survey is a standalone application using a smaller, static data set. The participants said that they would like to see PaperCube interface with a live data set. PaperCube is web-based and retrieves meta data using



JSON (JSON, 2009). As a result, it would be easy to integrate directly with an publicly available digital library such as CiteSeer<sup>x</sup> (Pennsylvania State University, 2009). The interaction model with a live digital library could manifest itself in two ways.

First, we could add more search capabilities which would allow PaperCube to search the digital library directly, and have it exist as a standalone service. The benefit of this approach would be that PaperCube is already developed to be standalone and only a data interface to the digital library would have to be written. However, rich searching capabilities of existing web-based digital libraries would have to be duplicated inside PaperCube which might distract from its central goal of focusing on the visualization of bibliographic meta data. Adding additional features to the application to make it a full interface to a digital library would most likely be a distraction.

Second, we could make PaperCube more lightweight and incorporate it into an existing web-based digital library as an additional feature. The idea is that when a user views a paper in the digital library, a link on the page would allow the user to view that paper in PaperCube. Then, from PaperCube, the user can navigate and view papers and authors like in the existing application.

The second interaction model with a digital library seems like it would be most desirable and effective. PaperCube was designed and developed with the goal to augment existing digital libraries through the use of visualizations, not replace them.

Another possible extension to PaperCube would be to expand beyond viewing papers and authors and explore the visualization of other information with relationship links. At a demo of PaperCube, the suggestion to use the visualization framework to show the relationships between knowledge base articles was brought up. Also, other information could be visualized quite easily, especially using the NodeGraph class developed as part of PaperCube. As a result, PaperCube and the NodeGraph class were released under the MIT open source license (PaperCube, 2009b).

## 9. Conclusion

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

The user study validated that the set of views developed for PaperCube achieved the goal of effectively augmenting digital libraries. Participants unanimously noted that PaperCube would fit well within the framework of an existing digital library service and that if it were available, they would use it. Although the data set used focused on papers published in the engineering and scientific spaces, participants from other disciplines including law thought that with the right data, PaperCube could improve their daily workflow.

Participants found that PaperCube helped reduce the cognitive load of a researcher by making it simple to really focus on a given paper or author, and made it easy to find the most relevant related publications. The views gave the user ability to change the amount of data visualized in real time through a set of parameters that could be altered through a set of sliders. Especially in data dense views, this customizability made it simple to determine what papers or authors are important by altering factors that indicate significance. Therefore, instead of trying to automatically determine what is relevant and eliminating the user from the equation, PaperCube gives the user control. Finally, it was shown that PaperCube is a “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] Bergström, P. (2009). Augmenting digital libraries using web-based visualizations. Master's thesis, Santa Clara University, Department of Computer Engineering.
- [2] Bergström, P., Atkinson, D. C. (2009). Digital library visualization using SproutCore and SVG. In *Proceedings of the Seventh International Conference on Scalable Vector Graphics*, Mountain View, CA.
- [3] Bergström, P., Whitehead, E. J. (2006). 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.
- [4] Ferraiolo, J., Jun, F., Jackson, D. (2003). Scalable Vector Graphics (SVG) 1.1 specification, W3C recommendation. Technical report, World Wide Web Consortium.
- [5] Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 17(4), 16–23.
- [6] Giles, C. L., Bollacker, K. D., Lawrence, S. (1998). CiteSeer: An automatic citation indexing system. In *DL '98: Proceedings of the Third ACM Conference on Digital Libraries*, (pp. 89–98)., New York, NY: ACM.
- [7] Hickson, I. (2009). HTML5 - Draft recommendation. Technical report, Web Hypertext Application Technology Working Group.
- [8] Hopkins, D. (1991). The design and implementation of pie menus. *Dr. Dobbs's Journal*, 16(12), 16–26.
- [9] Johnson, B., Shneiderman, B. (1991). Treemaps: A spacefilling approach to the visualization of hierarchical information structures. In *VIS '91: Proceedings of the 2nd Conference on Visualization '91*, (pp. 284–291)., Los Alamitos, CA: IEEE Computer Society Press.
- [10] JSON (2009). JSON (JavaScript Object Notation). Retrieved from <http://www.json.org>.
- [11] Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S. (2002). The open archives initiative protocol for metadata harvesting. Technical report, Open Archives Initiative.
- [12] Lamping, J., Rao, R., Piroli, P. (1995). A focus+context technique based on hyperbolic geometry for visualizing large hierarchies. In *CHI '95: Proceedings of the SIGCHI conference on Human factors in computing systems*, (pp. 401–408)., New York, NY: ACM Press/Addison-Wesley Publishing Co.
- [13] Lee, B., Czerwinski, M., Robertson, G., Bederson, B. B. (2005). Understanding research trends in conferences using paperlens. In *CHI '05: CHI '05 Extended Abstracts on Human*

*Factors in Computing Systems*, (pp. 1969–1972)., New York, NY. ACM.

[14] Lynch, C. (2005). Where do we go from here? The next decade for digital libraries. *D-Lib Magazine*, 11(78).

[15] Mackinlay, J. D., Rao, R., Card, S. K. (1995). An organic user interface for searching citation links. In *CHI '95: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, (pp. 67–73)., New York, NY. ACM Press/ Addison-Wesley Publishing Co.

[16] Modjeska, D., Tzerpos, V., Faloutsos, P., Faloutsos, M. (1996). BIVTECI: A bibliographic visualization tool. In *CASCON '96: Proceedings of the 1996 Conference of the Centre for Advanced Studies on Collaborative Research*. IBM Press.

[17] MySQL (2009). MySQL 5.1 reference manual : 11.8.2 Boolean fulltext searches. Retrieved from <http://dev.mysql.com/doc/refman/5.1/en/fulltext-boolean.html>.

[18] National Science Board (2008). *Science and Engineering Indicators 2008*, volume 1. Washington, U.S. Govt. Print. Off.

[19] PaperCube (2009a). Introduction to PaperCube. Retrieved from <http://vimeo.com/3323956>.

[20] PaperCube (2009b). The PaperCube open source project. Retrieved from <http://github.com/pbergstr/PaperCube>.

[21] Pennsylvania State University (2009). CiteSeer\*: Scientific Literature Digital Library and Search Engine. Retrieved from <http://citeseerx.ist.psu.edu>.

[22] Robertson, G. G. , Mackinlay, J. D. (1993). The document lens. In *UIST '93: Proceedings of the 6th Annual ACM Symposium on User Interface Software and Technology*, (pp. 101–108)., New York, NY. ACM.

[23] Shen, Z., Ogawa, M., Teoh, S. T., & Ma, K.-L. (2006). BiblioViz: A system for visualizing bibliography information. In *APVis '06: Proceedings of the 2006 Asia-Pacific Symposium on Information Visualisation*, (pp. 93–102)., Darlinghurst, Australia. Australian Computer Society, Inc.

[24] Sprout Systems Inc. (2009). SproutCore: A JavaScript framework. Retrieved from <http://www.sproutcore.com>.