Building for Users not for Experts: Designing a Visualization of the Literature Domain

Sarah Faisal, Paul Cairns, Ann Blandford
University College London Interaction Centre (UCLIC)
Remax House, 31/32 Alfred Place
London WC1 E7DP, UK
+44 (0) 207 679 5225
{s.faisal, p.cairns, a.blandford} @cs.ucl.ac.uk

Abstract

As researchers we are constantly working with academic literature. Literature data is growing exponentially. Interacting with this huge amount of information has been a challenge for the field of HCI for years. The goal is to assist users in making sense of this information by producing usable designs. Information Visualization (InfoVis) augments users' cognition when interacting with complex data structures. Although the use of InfoVis as a tool for representing literature data is not new, we have found that most of the existing Literature Knowledge Domain Visualizations (LKDViz) target specialists known as domain analysts who study publication patterns. Our goal is to design a LKDViz tool for academic users. Due to the diversity of academic literature users we captured their experiences and used it as the main source for our design. Interestingly, this method generated design criteria that have not been applied in most of the academic literature visualizations.

Keywords: Information Visualization, Literature Knowledge Domain, User Experience.

1. Introduction

As researchers, we are constantly surrounded with academic literature. Making sense of this exponentially growing information is a constant challenge. Lots of work in the field of HCI is being conducted to improve the design of such complex systems, more specifically when it comes to digital library design, for example [2].

Information Visualization (InfoVis) applications are powerful in that they augment cognition by using people's natural visual perceptions. By the simple action of looking at visually represented data the user can, at a glance, extract information that would have taken a long time to accomplish using, for example, text-based representations. The use of InfoVis systems as tools to manage literature data is not new. Literature Knowledge Domain Visualizations (LKDVis) exist; however these systems target only expert domain analysts who are interested in the evolution and publication patterns of scientific domains [4].

People are different. Designing for a broad user-base and not the expert increases the variance between users, so how do we cater for these differences? In other words, how do we design InfoVis systems for general academic users and not for experts?

The main challenge behind this work is based around the fact that working with literature is an experience that differs from one researcher to the next. In this paper, we identify a systematic approach for designing academic literature visualizations which begins with extracting users' experiences whilst interacting with their Literature Knowledge Domains (LKD). From this a descriptive theory was generated. Based on this theory we developed a design that would satisfy the diverse users and their various goals, as described bellow.

2. The literature domain: an experience

Following are three scenarios of researchers of varying experiences interacting with their literature data. These quotations are extracted from the qualitative study we conducted [5].

First Scenario: First Year PhD student

"finding papers has been a real problem in terms that I just found too many because you read one and then from that you read three more and so you get a pyramid effect in it and you end up collecting so many papers that it is just impossible to read" – making sense of the domain by browsing through the literature

Second Scenario: An Experienced Researcher (more than eight years of experience)

"I would look more at the fine details than I would have done when I started out but that is because I didn't know how much more important they would be whereas now they are extremely relevant because they are the only thing that kind of separate one paper by someone from another" – critical reviewing of the literature

Third Scenario: An Expert Researcher (more than 15 years)

"Um, when I go out to look for stuff to read I must admit its either because I have realized I have got a particular question…or most commonly actually is because I want to write a paper and I am aware that I don't know quite enough about what other people have done before I have done my own work and so I have to kind of back fill in on the literature" – look for specific literature

Researchers use academic literature differently at various stages of their careers. The means with which researchers interact with literature depends on their experience, knowledge and goals. We can see that interacting with the literature domain is an experience that differs from one person to the next.

3. InfoVis and the literature knowledge domain

Even though researchers' literature experiences differ, their interaction with the literature domain is an inseparable part of their careers. The academic literature data is complex with lots of interrelated dependencies, such as citation and publication links that can be difficult to follow and make sense of. InfoVis is an ideal means for representing this information since it augments users' cognition. By looking at a visual representation of the data users can easily make inferences and discover interesting patterns and relationships that would have been difficult to identify. The use of InfoVis as a technique for visualizing academic data is not new. Academic literature has been presented by many visualization tools. We identified two classes of academic literature visualization tools: Knowledge Domain visualization tools (KDViz), these tools paint generalized pictures of academic domains, and Information Retrieval tools (IR).

As indicated, interacting with academic literature is a subjective experience which differs from one user to another. In order to cater for such subjectivity we identified three factors that are essential for designing academic literature visualizations with a wide user population. These factors are:

- Interactivity
- Requirements gathering
- Usability studies

Interactivity

Interactivity is one of the core criteria since it allows for the realization of subjective user experiences [6]. Through the interactivity of the tool, the users are able to achieve their varying goals by manipulating the visualization. Users can address these goals through the use of onscreen widgets; for example: panning the visualization through the use of a slider; or by interacting directly via objects of interest; for example: selecting the object of interest by clicking on it via the mouse.

Requirements

The needs of academic literature users vary greatly, as we have indicated. As a result, literature visualization tools must reflect the varying user experiences and needs. We strongly argue that user requirements must be gathered. Requirement gathering is standard practice in HCI. However, when it comes to designing academic literature visualizations this is not the case, as will be discussed next. Portraying concepts that do not reflect user needs will make the visualization tool uninteresting to the user population and hence little knowledge will be gained.

Usability

The tool must be usable. This ranges from the clarity of the visual cues and their ability to portray the domain concepts, to the usability of the tool in general.

In this section we will discuss the work done in the areas of KDViz and IR from these three perspectives.

3.1. KDViz: Knowledge Domain Visualizations

Academic literature visualization is a well established research area that is part of a growing field known as Knowledge Domain Visualizations (KDViz) [3]. These tools analyse the data via computationally intensive algorithms. The generated visualizations satisfy specific goals of an expert population of users. In fact most of the work done in KDViz that represents literature data is designed for specialized expert users known as domain analysts. Domain analysts have a very specific goal which is to analyse the development and evolution of scientific domains. Users of these tools engage in tasks such as determining the varying trends in citation and co-citation networks [4]. Generally speaking, domain analysts seek a representation of the whole represented in global domain visualization and not

of the specific domain details reflecting individual entity details.

Interactivity

These visualizations are not interactive in the sense that we are looking for in an InfoVis tool. Users are only able to manipulate the visualization through limited actions such as zooming and panning the interface. This is due to the goals of these tools in addition to the high computational power needed to generate these visual representations.

Requirements

Although there has been no evidence in the literature of any requirements gathering, users of these visualizations are expert users, which in most cases reflect the developers themselves. Hence, we might deduce that the requirements with which they build the tools fit their needs.

Usability

Another important factor is the usability of the tools. From reviewing KDViz literature most of these tools have not been user tested, hence we can not be sure of their usability.

3.2. IR visualization tools

Literature IR visualizations represent the second category of literature visualizations. The tools in general assist users in searching for and retrieving literature data. More precisely the tools assist users in better understanding and browsing search results. The tools concentrate more on the details of the literature domain rather than the 'whole'. Unlike KDViz, literature per se is not what the IR developers concentrate on. Literature in this case is merely a domain that they use to reveal the ability of the developed visualisation tools, meaning that there is not a tight coupling between the domain and the tool. The same visualization can be used for data that is not necessarily related to academic literature, for example: Butterfly [9] and GRIDL [11].

From reviewing literature in this category the notion of the 'users' was evident, which was not the case in KDViz literature. In addition, usability studies hold an important part in the development process of these tools.

Interactivity

These visualization tools are very interactive. Users manipulate these visualizations by clicking on on-screen widget, such as: buttons, sliders, etc. User can perform actions such as: searching for entities of interest where they can identify an author or a paper of interest, they can view specific citation links between entities, and visualize the same data in multiple forms. What is important here is that the user can interact with the specific entities unlike with LKDViz where this type of interaction is limited.

Requirements

Although in certain systems, such as Envision [10], requirement gathering is performed, this step is not an essential step in the development process of these visualizations since, as indicated, the literature as a domain is not the goal. Hence, it is not always the case that the requirements around which these tools were built fit the needs of the academic literature users.

Usability

Most of the systems developed under this category go through a series of usability studies. However, we noticed from reviewing literature related to the development of these tools that these studies concentrate on the usability of the features of the tools rather than whether these tools fit within the literature domain's requirements per se.

To summarize the above, we state that the literature experience is a subjective one. Existing academic literature visualization tools do not cater for the users' literature experience diversity. Since researchers' literature experiences differ we need to design for user diversity and not for the domain expert. This is where our challenge lies, since we need to identify a design that satisfies these users. To do so we need to understand how researchers work with their literature domains, in other words capture the user experiences.

4. Qualitative understanding of the literature domain

In order to capture user experience of the LKD we conducted a qualitative study [5]. The aim of the study was to understand how researchers, of differing experiences and knowledge, make sense of the huge literature domains at various stages of their careers. Following is a summary of this qualitative study.

4.1. The study

During the study semi-structured interviewing was used as the main information gathering technique due to its flexibility and ability to reveal concepts that were not previously evident. The interviews were conducted with researchers of differing experiences ranging from a first year PhD student to a Professor in the fields of HCI and psychology. Each interview was fully transcribed and analysed using the Grounded Theory methodology [12]. Categories and concepts were extracted through open and axial coding. The analysis of each interview guided the formation of the next interview. This continued until a saturation point was reached. The saturation point was reached after a total of eight interviews. This quick reaching of the saturation point is unusual for this type of study, but was quite interesting. We think that it might be an indicator of the concreteness of the literature domain.

The interviews revealed the subjectivity of the literature domain. The participants' literature knowledge depended heavily on their background, knowledge and goals at the time of interacting with their LKD. From these interviews a descriptive theory [5] was generated.
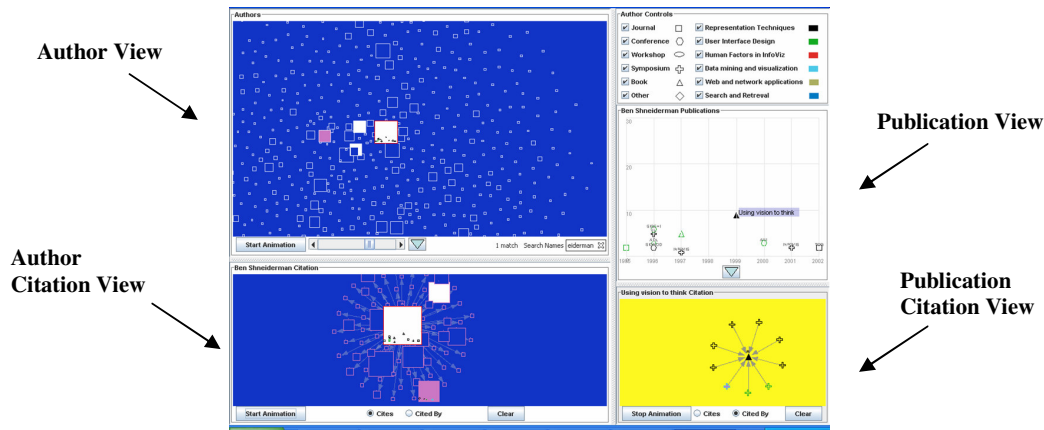
Figure 1. LKDVis Primary Prototype

4.2. The LKD descriptive theory

The descriptive theory describes how researchers, at least across the psychology and HCI domains, make sense and gain knowledge of their literature domains. It can be summarized as follows:

The sense-making activity of the academic literature domain starts with getting to know who the members of the community are, since they are the creators of the knowledge P2: *"this [community] is where I look for literature"*. This knowledge is portrayed in the form of literature which is produced by the members. It is through the interaction with such literatures that ideas are generated and personal knowledge of the domain is gained. The ideas that the papers reflect are what is important to the researchers, P6: *"… I think I would go for ideas… what it means actually it is not the paper but the ideas"*. The sense-making process of ones' literature domain is based around users' engagement with a set of activities such as: searching for specific authors, ideas or papers, answering specific questions, and following citation trails. In most cases the gained knowledge is not intentional but is gained as researchers interact with the domain. For example: during one of the interviews when the participant was asked how s/he determined who the influential author in their field was the participant said: *"… those people that are really important you can't miss the name because they are always cited and they always come up"*.

Following from the descriptive theory we designed and implemented a visualization prototype of academic literature. The design of our visualization tool was inspired from our descriptive theory. As a result, when we explain the design of the tool we will refer back to the theory by giving examples of some of the quotations our participants gave during the course of the interviews.

5. From theory to design: developing LKDViz

The system rationale is based upon the individualization of the experience, as revealed by the descriptive theory. Hence, the goal is to give the user the freedom to explore and manipulate the visualization through multiple interactive activities. In addition, users need to be given the ability to personalize their experiences, and by that we do not mean simply changing the visual cues such as color or shape in order to fit in their aesthetic pleasures. We mean that users need to affect the semantics of the visualization. Furthermore, it is crucial for researchers to visualize themselves as part of their research communities and see how their research evolved within their communities. All of these interesting findings differ from existing academic literature visualization tools.

We developed a preliminary prototype (Figure 1) of a literature domain. The data we visualized was the dataset used for the InfoVis'04 contest which includes the complete metadata of 8 years for all InfoVis conference papers and references from 1995 to 2002[8]. The design decisions were based exclusively on the results of the descriptive theory.

In the above figure, we see the entire domain represented on the left. The domain is represented by its authors, the user in this example clicks on an author for example, "Ben Shneiderman" as seen in the figure. Shneiderman's papers are displayed on the right hand side in the articles view. Detailed description of the layout of
our prototype and explanation of the different views in addition to the different functionalities of the tool is explained next.

5.1. Note on design features

The interviews revealed a concrete set of data elements which include: author, article, source (e.g. journal, conference, workshop, book, etc), in addition to relationships such as citation, authoring, and collaboration. Lots of guidelines exist in assisting designers in determining which visual attributes best fit the intended meanings. In our prototype, for example, we relied on [1] and [13] as a basis of our visual design decisions.

5.2. Data layout

The theory revealed that the authors, in other words the members of the community, are at the center of all domain sense-making activities. This is due to the fact that they are the producers of knowledge. Ultimately, papers are written by them and hence they cannot be separated, e.g. P2:*"It is hard to separate that [articles] from authors, cause ultimately they were written by authors"*. In addition, during the course of the interviews it was observed that the interviewees tended to remember the name of the influential authors better than those of influential articles. As a result, we decided to make the authors view the main visualization view with which the user interacts with the tool, as we will explain next.

From the interviewees' perspective the authors are made out of a group of papers, giving the author a particular status. Hence, each author is represented on the screen as a square grouping his/her papers (see Figure 2).
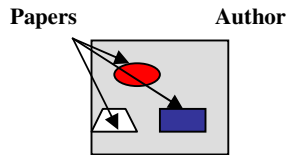
Figure 2. An Author as a Group of Papers

The color of the paper represents the interest and the shape of the paper represents the source (e.g. journal, conference, etc). Merely by looking at the author the user can identify the interests of the author and the amount of publications the author produced, since the size of the author depends on the number of publications.

In order to give the user the freedom to manipulate both the authors, publications and associated citation relationships, in addition to keeping the identified author-paper relationship, we decided to divide the interface into 4 views (see Figure 1). The main view is the authors view, which is the bigger view on the top left. And the details view is the publication view seen on the top right. As the user selects an author, all the author's papers are displayed on the paper view. The bottom left view represents the citation information between authors, whereas the bottom right view represents the citation information between publications.

The layout of the authors view takes the form of a force-directed layout graph, where nodes by default repel each other and edges act as springs bringing related nodes closer to each other. We used the Prefuse visualization tool- kit [7] to develop this visualization. The author graph was based on the collaboration relationship between authors. We chose collaboration as the clustering parameter because of the results of the qualitative study we conducted where participants indicated that authors relate through collaboration links. For example, during the course of a particular interview when one participant was asked about how authors relate, she indicated: *"they are co-authors on the same papers …".*

The details view which represents the publication view, represented on the right in Figure 1, is organized in an x-y grid layout, where the x-axis represents the year of publication and the y-axis represents the citation ratio of the paper. The more a paper was cited, the higher it is located on the axis. Like the collaboration layout, this layout was decided upon based on the results of our qualitative study.

5.3. Interactive activities

The strength of interaction lies in its power to manipulate the visualization according to users' varying goals. From our descriptive theory we generalized a set of activities that, when engaged in, would enable users to gain more insight into their LKDs. These activities incorporate both an interactive task and visual feedback by the system revealing additional information.
Collaborations
In addition to the first author the participants indicated that there was a need to identify the other authors of the paper.

P5: "I always need to know the *second* and or *third author*".

In addition to the collaboration graph, our prototype supports the collaboration relationship via the combination of the authors and publication view. When the user clicks on a paper in the publication view the collaborators will become visually distinguishable on the authors view.

Citations
Following citation links is an essential part of the literature sense-making activity since it allows users to discover hidden information, for example: important authors or papers that the user was not aware of. E.g. P3: "I would use his [influential author] *paper* as a way of kind of *giving me pointers* to *other people* whose work I might look at".

Interestingly, the theory revealed that citation links that are of interest to the users are not just, as we might assume, between the papers, but also between authors. Hence, we included the authors citation view which represents the citation information between authors, and the publication citation view, which represents citation information between papers.

In addition, the participants indicated that there was a need to identify citation links between authors and the papers, for example to see all the authors that a paper cited. E.g. P1: "I also look at *people* who have cited the *paper*".

Users of our system can see the citation relationships between authors and between papers. For example, the user can see who cited this author by simply clicking on the object of interest which is either the author and inserting it into the corresponding citation view; as a result all the authors that have papers that cited the selected author are revealed. The same thing applies to the articles. In addition, the tool always relates the information that is displayed in the various views.

For, example when the user click on an author in the author citation view, all publications that are published by that author in the publication view and publication citation view become visually distinguishable. Same applies when clicking on a particular paper in either the publication, or publication citation view all authors that collaborated on that particular publication become visually distinguishable.

Personalization of Knowledge

It is essential to give the users enough freedom to customize the visualization by layering their personal views on top of the original visualization. It was revealed that when it came to certain concepts, there was some variance. For example, some interviewees defined influence as being related to the number of citings an author has got, whereas others defined it as an author or a paper that has been influential to their own ideas and thoughts. E.g. P4: *"I suppose when you say influential I consider it to be influential to my own ideas"*. It would be easy for the tool to calculate the number of citings an author has; however it would be impossible for the computer to determine user preference since it will depend entirely on each user. Therefore, we propose the idea of overlaying the visualization with the user's personal views, allowing the user to add thoughts and comments to the visualization entities. The means with which these personalization features will be implemented into the design of prototype is part of our future work.

In addition to personalizing the visualization, the qualitative study revealed the ultimate need for users who, as we indicated, are researchers, to visualize themselves and their work as part of the community. E.g. P2: "…you want to know how *your work relates* to others…" This will cater for a wider research audience, since, as the interviews revealed, as researchers become more and more expert in their fields the need to make sense of their literature domains decreases; however the need to understand how their research evolved and is embedded within the community becomes of importance.

6. Future work

We will design and explore the personalization features. Following, detailed user studies will be conducted on this prototype. Furthermore we will be altering the appearance of the on-screen widgets to determine which is preferred by the users. These studies will also test the findings of our descriptive theory.

7. Conclusion

InfoVis systems that represent academic literature data exist; however, most of them target expert users. When designing for experts, little diversity exists since they have very specific goals. It is when designing for diverse users that the challenge arises.

In order to address this challenge we captured the users' experiences with the goal of understanding how researchers work with and make sense of their literature domains. This assisted us in understanding the variance and designing for it. We developed a preliminary prototype that differed greatly from most LKDViz.

In this paper we shed light on a major problem facing academic literature visualization tools which is to assist users in making sense of and managing, large amounts of information by producing usable designs. To conclude we say that before we can design for a better user experience this experience should be well understood.

References

[1] Bertin, J. *Graphics and Graphics Information Processing*. Berlin 1981. Gruyter, W. being a translation of Bertin, J. *La Graphique et le Traitement Graphique de l'Information*, Paris, Flammarion 1977.

[2] Blandford, A., Keith, S., Connell, I. & Edwards, H. Analytical usability evaluation for Digital Libraries: a case study. *Proc. Joint Conference on Digital Libraries*, ACM/IEEE (2004).

[3] Borner, K., Chen, C. and Boyack, K. Visualizing Knowledge Domains. *Annual Review of Information Science and Technology* 37 (2003), 1-179.

[4] Chen, C. Searching for intellectual turning points: Progressive Knowledge Domain Visualization. *National Academy of Sciences of the United States of America* (PNAS), 101, 1 (2004), 5303-5310.

[5] Faisal, S., Cairns, P., and Blandford, A. Developing User Requirements for Visualizations of Literature Knowledge Domains. *Proc. Information Visualization 2006*, IEEE (2006).

[6] Faisal, S., Cairns, P., and Blandford, A. (2006) Subjective Information Visualizations. *V&I workshop on Combining Visualization and Interaction to Facilitate Scientific Exploration and Discovery*, British HCI.

[7] Heer, J., S.K. Card, J.A. Landay. prefuse: A Toolkit for Interactive Information Visualization. *Proc. CHI'05*.

[8] Ke, W., Borner, K. and Viswanath, L. InfoVis04 Contest MS-Access Database, Indiana University, School of Library and Information Science and School of Informatics, 2004. http://ella.slis.indiana.edu/~lviswana/iv04contest.mdb.

[9] Mackinlay, J. D., Rao, R. and Card, S. K. (1995) An organic user interface for searching citation links. *Proc CHI' 95*.

[10] Nowell L. T., France, R. K., Hix, D., Heath, L. S. and Fox, E. A. (1996) Visualizing search results: some alternatives to query-document similarity. Proc ACM SIGIR Conference on *Research and development in information retrieval*.

[11] Shneiderman, B., Feldman, D., Rose, A. and Ferré Grau, X. (2000) 'Visualizing digital library search results with categorical and hierarchical axes', In Proc ACM *conference on Digital libraries*.

[12] Strauss, A. and Corbin, J. M. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. SAGE Publications, 1998.

[13] Tufte, E.R. *Envisioning Information*. Graphics Press, Cheshire, CT, 1990.