

# Developing User Requirements for Visualizations of Literature Knowledge Domains

Sarah Faisal, Paul Cairns, Ann Blandford  
UCL Interaction Centre, University College London  
31-32 Alfred Place, London, WC1E 7DP, UK  
s.faisal@cs.ucl.ac.uk, p.cairns@ucl.ac.uk, a.blandford@ucl.ac.uk

## Abstract

*As researchers we are constantly working with our literature domains. Making sense and familiarizing ourselves with such domains is a cumbersome but necessary task that is part of every researcher's career, especially novices. Literature Knowledge Domain Visualizations exist which assist users in extracting and interpreting interesting patterns. However, these visualizations tend to be developed without reference to the particular needs of researchers or without addressing the specific challenges of the literature knowledge domain. This paper therefore targets the needs of researchers as they work with their literature. We conducted a qualitative study to understand how researchers make sense of their literature domains. Such information assisted us in coming up with a set of design considerations and guidelines which we believe will assist in building visualizations that are useful for the general research population.*

## 1 Introduction

Starting off as a novice researcher in a novel domain is a burdensome task, since one needs to get familiarized with the domain. Researchers have to go through huge amounts of literature to try and understand the structure and settings of the domains to which they belong. The field of Information Visualization (InfoVis) seems ideal for assisting in such a task since it provides the means for representing information on the screen using visual cues, which in turn reduces the effort needed to gain knowledge. For example, it would assist researchers in extracting patterns and building associations faster than when relying on standard textual hyperlink mechanisms. Using InfoVis as a means to solve such a problem is not new. Previous work has been done in knowledge domain visualization (KDViz) where literature knowledge domains have been represented visually. Such visualizations use computer generated algorithms that extract certain patterns and in turn represent them visually on the screen [4]. These visualizations however do not seem to have been built to target the particular needs of researchers.

We argue that working with literature forms an inseparable part of the experience of researchers and hence this

work approaches KDViz from the perspective of the researchers needs, goals and methods. The study described here explores the means and the tools that researchers use in order to make sense of their literature domains through a descriptive theory which sheds light on this phenomenon. It reveals the importance of users' subjective and personalized experiences when interacting with such a domain. To our knowledge this user perspective has not been previously highlighted. From such an analysis, a proposition is presented as to how such information can be used in creating literature domain visualizations revealing a set of design guidelines, one of which is interaction which supports users' subjective experience. This proposition and associated guidelines, we believe, will make such visualization useful for the general research community.

## 2 The literature domain

One of the major goals of Knowledge Domain Visualizations (KDViz) is the representation of knowledge by visually painting a picture of the scientific development and evolution of a domain. In other words, it attempts to give a general view of what is going on across a particular discipline over a period of time. The ideas these studies tend to represent may include for example: how popular a particular field is, how popular a particular author is, the evolution of a particular field, the diffusion of research topics, etc, all within a specific domain. These visualizations use specialized algorithms which in turn extract certain patterns and semantic similarities [2] out of the domain's data, and represent them visually on the screen. These patterns are then used to visually represent citation patterns, such as the evolution and significant contributions within a knowledge domain [3], where the entire citation domain is visualized. It is claimed that these visualizations assist audiences with special interests, such as domain analysts and experts, yet we have found no evidence of usability studies to prove such a claim. In addition, we see that there is more to this domain than what is being represented by these visualized patterns. Looking closer we can see that these visualizations are built on certain assumptions according to which such patterns are being extracted and represented. These assumptions are not based on re-

searchers' experiences and needs of such a domain. As indicated, the knowledge gained out of the literature domain is of interest to the general research community. Any researcher when entering a specific field goes through the primary step of familiarization in which they try and make sense of their literature domains. Since we are interested in building visualizations that target the general research population, we need to verify whether or not the assumptions used in building the existing literature domain visualization tools comply with researchers' literature sense making experiences. For example, one of these assumptions is: each publication in a field can be characterized by its list of keywords [1], however one of the researchers who we interviewed indicated that they do not trust keywords since in certain cases they do not really reflect what they are looking for in a document. This example indicated that there are additional or different criteria upon which different visualization patterns need to be presented. It is this type of knowledge that we are really interested in in this study.

In addition to the literature KDviz, there exists another category of research conducted in the area of visually representing literature data. However, the main intention behind these studies was the development of new interactive visual metaphoric solutions for representing large amounts of interrelated information. They were not really intended to assist users in making sense and working with their literature. In other words, the application domain was not in itself the problem, but how to represent large amounts of interrelated information in an understandable and usable way. Seminal examples of such work are Fairchild et al's SemNet [5] with its syntactic 3D representation, Mackinlay et al's [6] Butterfly with their organic interface, and Shneiderman et al's [7] GRIDL with the idea of categorical grouping.

From here we can safely say that it is not clear from all of these previous studies what information researchers truly need to visualize in the literature domain. Hence, this study has been devised to fill in the missing gaps which relate to researchers' literature knowledge domain requirements. It is our primary aim to understand researchers' experiences in the task of building and making sense of literature domains. From this understanding we will determine the literature domain information that needs to be portrayed in order for these visualizations to be useful by the general research population in addition to the means with which this can be achieved.

### **3 The study**

The study conducted was qualitative. It aimed at gaining a general understanding of the problem of learning about a new literature domain from a user's perspective. Interviews were conducted and analysed using grounded theory [8]; from this descriptive theory general design guidelines are proposed.

#### **3.1 Statement of purpose and study scope**

The main goal of this study is to derive a high level descriptive theory across both the psychology and the human computer interaction (HCI) fields in relation to understanding the different ways in which researchers make sense of, work with and perceive their literature domains. It intends to gather enough information to determine the different literature entities and tools that researchers work with in addition to the knowledge they gain in relation to their literature domains. From this descriptive theory, the basic requirements for a literature domain visualization tool will be determined.

#### **3.2 Method and setting**

Researchers are the main audience this study targets since they are the primary users of the literature domain. In order to get the intended high level understanding of the problem at hand, information needs to be gathered from researchers of varying experience and knowledge of a domain in which they are working or proposing to work. This study focuses on researchers from the domains of Psychology and HCI because they were available at the time of the study. Due to the type of information needed from the study, semi-structured interviewing was used as the data gathering tool. Since it is a fairly open and flexible framework it provides the opportunity to learn additional relevant information not previously known. In total, eight interviews were conducted with researchers of varying experiences (ranging from an experienced professor to a first year PhD student) in the fields of psychology and HCI. The questions asked emerged from the conversations held during the interviews. Outcomes that arose during the analysis of the gathered data guided the decision as to which subject to interview next, which depended greatly on the subject's previous research experiences. The analysis of these interviews was performed using the grounded theory methodology, which is an inductive theory development methodology. It assists in analysing the data from a higher-level case perspective which assists in identifying categories and concepts through open coding and relating these codes using axial coding, and finally determining a single storyline around which everything else is draped through selective coding. The process of gathering the data, analysing it and so on was continued until a saturation point was reached after a total of eight interviews. This rapid attainment of saturation is unusual but may reflect the very concrete attributes of what defines literature, and the broad agreement of how to approach literature in research, at least across the subject areas of the interviewees, .

#### **4 The overall literature domain sense-making story**

Grounded theory assisted in developing a descriptive theory from which the general literature domain sense-

making underlying story was revealed. As the details of forming a grounded theory can be quite complex, it is useful to describe here the overall story of the theory before giving the fuller details that make up the basis for the theory.

Researchers are constantly working and interacting with their literature domains. The information about a domain is strongly related to literature data such as the authors and their associated publications. Thus, a domain is quite simple in relation to the entities it incorporates yet a vast amount of complicated interrelated knowledge can be gained. This study reveals that the sense making process of one's literature domain starts with trying to determine who the members of the community are and their associated research interests. The community can be seen as a globe that surrounds its associated entities which represents the members, their interests and the literature they produce. We identified that it is the community that forms the primary source of the literature domain sense making process.

Various research interests exist within a community, each embedded with its members. Members of the community collaborate with one another to produce knowledge, which is reflected by the literature itself. It is work that goes into these pieces of literature that results in the advancement of knowledge, and hence the development and evolution of the specific research domain. The process of understanding one's literature domain results in gaining vast amounts of knowledge. Some of this knowledge is quite explicit and direct, such as who collaborated with whom on a piece of work, whereas some is more implicit and subjective such as who is influential in a particular domain, or what piece of work or idea changed the course of a field's development, etc.

The understanding of one's literature domain is one of the key phases researchers go through as part of their careers. It is through this understanding that they get familiarized with the research domain to which they belong, gain knowledge of seminal work in the area, and get ideas for their own research. It is this understanding that assists them in participating in the overall advancement of research.

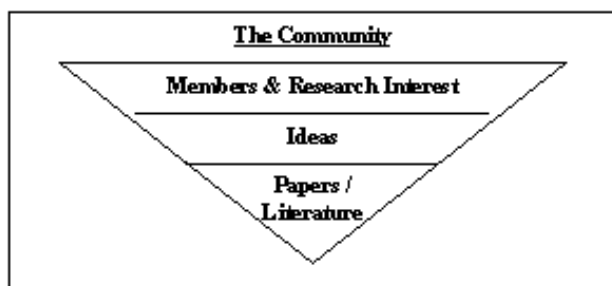


Figure 1: Understanding of One's Literature Domain

Sub7: "...I think that this is quite the whole point in a way, we are trying to **build up some picture** of um you know an understanding of whatever our area is".

We can say that the understanding and sense making process of one's literature domain starts with knowing who the members of the community are, in addition to their associated research interests (Figure 1). More importantly, it relates to the ideas that researchers generate of their domains. The papers, in other words pieces of literature, represents the lower level concrete entities that researchers interact with in order to gain information which eventually leads to gaining knowledge of their literature domains.

## 5 Analysis

Having given a general feel for the theory of the literature domain sense-making process, we turn to the detailed findings of the interview analysis. Grounded theory allowed for a number of categories and associated relationships to be identified through an iterative process of analysis using open, axial and selective coding. These results are further interpreted to reveal a close relationship with the literature domain understanding process.

### 5.1 Categories and concepts

Four main categories were identified during the open and axial coding processes which reflect the interviewees' subjective experiences while working with their literature domains. These are:

- The research community
- Literature
- Influence
- Evolution of a discipline

We discuss each of these separately as they form the basis for the complete theory.

**The research community** represents the central category around which the sense-making process begins. From a higher level perspective, it reflects a social activity where people with similar interests communicate to share ideas and knowledge. A researcher's community varies along its dimensional scope from global to local. Global reflects the outside community to which the researcher belongs. Contact with members of such a community is done occasionally, when meeting in conferences, via e-mail or through telephone conversations. On the other hand, the local community represents the immediate community which surrounds the researcher on a daily basis, for example, members of his or her research department. One's research community is not location specific instead it is specific to the interest it incorporates, which is shared by members of various locations. The community is made out of members who form the main entity around which the community exists. It is these members, the people, which participate in expanding the general body of knowledge. Such is

achieved through the ideas they possess which is reflected through their work, in other words their publications. It is these ideas that form a valuable source of information.

Sub2: *“this [community] is **where I look for literature**”*.

**Literature** represents the concrete piece of information from which knowledge is gained. It is through literature that knowledge is transferred and communities are formed. However, literature in this sense does not just represent the actual physical papers, but also the ideas which these papers reflect. These ideas are documented within papers, and are created from researchers' subjective views. It is these ideas that researchers generate in the sense making process that are of interest more than the actual physical papers per se.

Sub3: *“...there'll be a core body of people who are aligned with particular kinds of **ideas**”*.

Sub6: *“...I think I would go for **ideas**...what it means actually it is **not the paper but the ideas**”*.

Papers form the concrete entities from which knowledge of the literature domain is gained. These entities can be summarised as follows: authors' names, topic, keywords, title, abstract, introduction, conclusion, source and publication year, where most of this information is found in the paper's citation string. Even though authors and papers are two different entities, it is not possible to separate them.

Sub2: *“It is hard to separate that [articles] from authors, 'cause ultimately they **were written by authors**”*.

The distinction between both was not made apparent unless subjects were explicitly asked. For example:

1. Subjects tended to mix up the pronouns when referring to authors and/or papers.

Sub2: *“**I look at who** is out there doing that sort of work, **I look at those papers**”*.

There seems to be some sort of an association between author “*who*” and paper, as if they were equivalent.

2. Subjects associated authors with the citing task rather than the paper itself.

Sub1: *“**I also look at people** who have cited the paper”*.

In this example it is interesting the way the subject uses the word “*look at people*” and not “*look at papers*”, where the actual action involves explicitly looking at the papers.

In addition, during the interviews subjects always seemed to remember the names of the authors that they were referring to in their examples. However, when talking about a paper they seemed to either refer to it as “the paper written by ...” or they referred to it by explaining the ideas that the paper presented, again reemphasising the notion of ideas.

Information regarding the literature domain is gathered through researchers' engagement in certain activities which satisfy some of the goals they might have at the time.

The activities may include interacting with *concrete entities* such as: looking for authors, following citation links, looking at authors' collaborators' work, and identifying general ideas, topics and keywords. Hence, we can say that it is the low level paper entities that the researchers engage with in the sense making process, and it is the *ideas* that they generate out of such an engagement that causes knowledge to be gained.

**Influence** is related to the *ideas members* are producing. Some of these ideas are more influential than others. Influence is something that causes an *impact*. An *impact* is characterized by something that causes a change in the way people, members of the community, think and work. Whether such an impact is received with agreement or disagreement, it is still an indicator of influence.

Sub3: *“I could name you papers that have been influential ... in terms of **changing thinking** in a particular area”*.

Sub5: *“... you have to read what they [influential authors] are doing **even if you don't agree**”*

The changes that such an *impact* may cause may be on a *global* or a *personal* level. The *global* level indicates that the changes that occur will be reflected on the domain in general. On the other hand, the *local* level indicates that the changes may also occur on the *personal* level, affecting the actual researcher and his or her ideas and work and not the global community.

Sub3: *“...there have been papers that have been influential ...actually changed the way **I have thought of my work**”*.

Sub4: *“I suppose when you say influential I consider it to be **influential to my own ideas**”*.

It is crucial to point out that it is quite impossible to put an objective explicit measure on the amount of impact or change caused by a person, since this characterization is very subjective, even though there are certain concrete factors that may be taken into account such as, the number of citations an author gets or the status of the publication.

Sub3: *“I think **they [influential authors] are cited a lot**, but you get to hear about them for other reasons”*.

In contrast to the influential author, it is very rare to find a paper that is on its own influential. In most cases it is the work which is presented through a group of papers that is influential, rather than a single paper.

Sub5: *“...It is not often that one article always comes up”*.

Sub6: *“...I think I would go for **ideas**...what it means actually it is **not the paper but the ideas**”*.

Therefore, it can be concluded that influence is a quite subjective matter which is affected by the views that different people may have. Hence, it is something that cannot be readily measured.

**The evolution of a discipline** is something that is related to a change in the status quo. It can be defined as the way in which a discipline progresses or evolves through time. It is normally caused by radical *changes* in influential people's ideas over a period of *time*. Evolution occurs when such ideas are met with general acceptance by members of

the community.

Sub5: "...evolution is kind of a development or **how a discipline developed from other disciplines**".

Sub7: "...there is that thing of collaboration so that you work together and then perhaps people from **different disciplines can perhaps contribute to one another's work and it'll develop that way**".

The understanding of the evolution of one's literature domain did not seem to be of major interest to the interviewed researchers. Some interviewees expressed interest in the topic of evolution, whereas others did not.

Sub4: "...it is a bit **beyond my scope**".

Sub5: "...I have always **been interested** in linking disciplines together".

Sub7: "...I suppose it is about **what people find interesting** and, you know, decide to invest their time in".

## 5.2 Interpretation

By interpreting these interrelated categories further, they can be broken down into four different parts which relate to the process in which the literature domain understanding occurs. These parts are: the concrete data, the activity, the perceived information and the knowledge gain. The knowledge gain reflects the process by which the knowledge and understanding of the domain occurs.



Figure 2: Breaking the Literature Domain Understanding

The concrete data reflects the pure factual data from which the researchers make sense of their literature domains. It represents the entities they work with throughout the understanding and knowledge gain process. It includes both the *community* and the *literature* categories and their associated subcategories. Throughout the sense making process the main analysis revealed that there is always an activity with which researchers are involved. As a result of people's engagement with such an activity sense making occurs. Such activities as indicated are involved with interacting with the concrete literature entities. As a result of interacting with such data additional information may be

deduced, such as forming subjective views reflecting influence and discipline evolution. Such perception is derived from the concrete data in addition to an individual's point of view and interpretation, hence forming a subjective perception and not facts.

The learning and knowledge gain of the domain is the major goal behind interacting with one's literature domain. When interacting with the concrete data a certain amount of information is perceived, and therefore knowledge about the domain is gained. The learning of one's literature domain is a slow and in some cases unintentional process that happens gradually with time, as soon as the researcher's career begins, and continues throughout the researcher's career. Sub1: "...so it was a **slow process** of learning where to look for these things, reading them and just getting interested".

Sub6: "...**no, it just happened**...This is very dangerous because I may miss out".

Sub7: "**I didn't set out to do that deliberately**".

From this interpretation, it can be concluded that the understanding and sense making of one's literature domain starts with the concrete data. By interacting with the data through a number of activities, knowledge of the domain is gained and information is perceived. This knowledge may differ from one researcher to another, depending on their needs and experience at the time. This perception is also considered as knowledge being gained. Therefore, it is safe to say that working with the literature domain is an experience that forms a major part in any researcher's career.

## 6 From theory to design

This descriptive theory revealed the information researchers needed and gained as part of their literature domain sense making process, in addition to the means and tools with which this was achieved. The findings of this theory can be interpreted in the form of general design guidelines which in turn form the general requirements around which literature domain visualization tools need to be designed and built. Three main design guidelines are identified: the visual representation, the subjective experience and the personalization qualities.

### 6.1 Visual representation

In addition to the need to represent the basic literature concrete entities, all relationships must be visualized as well. The study revealed that in certain cases the distinction between authors and their associated papers was not made evident. Hence, the visualization tool must be as flexible as possible in that it should allow users to see both the members, papers and their associated relationships, and not one or the other as exists in many existing LKD visualization tools in which they use single units of analysis which is typically either author e.g. [3] or

article e.g. [2].

## 6.2 Interaction: subjective experience

The interviews revealed that working with one's literature domain is quite a subjective experience, hence the knowledge gained from such an experience differs from one researcher to the other. Such knowledge depends immensely on the researcher's background, current knowledge and goals at the time. The theory revealed that the sense making knowledge gain was achieved as part of executing certain activities. It is the engagement with these specific activities that leads researchers to gaining knowledge of their literature domains. Hence, it can safely be stated that researchers working in similar research domains do not all make sense of their literature domain in the same manner even though they all work with similar concrete entities. As a result, the visualization tool should support such subjectivity of the experience by supporting users' varying goals and motives. The only foreseeable manner in which the tool would allow for such a subjective experience is for the tool to be *interactive*. Hence subjective experience equates with users interacting with the tool. In other words, we regard manipulation as essential since static visualizations only allow for a limited and concrete amount of information and knowledge to be gained. This emphasises the necessity of building a system which supports various interactive activities which in turn would enrich the experience.

## 6.3 Personalization: the introduction of ideas

Another interesting and crucial finding of the theory, which also affects the design guidelines of the LKD visualization tool, was the introduction of the notion of *ideas*. As indicated, researchers when making sense of their literature domains tend to be interested in the ideas that the members or the papers express more than the concrete papers *per se*. *Ideas* represent the major link that ties up the literature domain understanding process (Figure 1). Due to its importance and criticality within the knowledge gain process, it must be supported as part of the visualization tool.

However, the challenge lies in that these *ideas* are quite personal and do not pre-exist as part of the concrete literature entities. For example, the same papers might be read by two researchers each with a different interest in mind, hence different ideas would be generated. The descriptive theory revealed information with regard to the notion of *ideas*. For example, researchers associated authors to ideas more than to papers, where the same idea may be reflected in various papers which may be written by various authors, etc. Up to this point no knowledge visualization tool supports such a concept in which personalization is supported and represented as part of the tool. Adding such a concept into a visualization tool is quite challenging with

unknown consequences which might affect the visualization experience as a whole. Future work intends to tackle such a challenge and explore its effects on the visualization experience.

## Conclusion

We have presented a different perspective to the literature knowledge domain visualization tools. We have shown that in order to visually represent knowledge, personalization is crucial. Users must be able to incorporate their individualities into the visualization experience. Hence, we claim that static visualizations or visualizations that allow for users to merely fly through the visual representations are not the answer. We have revealed that these visualizations should be built around a set of activities which incorporate users' literature domain experiences. We claim that having an interactive visualization would be the answer. By interactive we mean visualizations that would allow for users to gain subjective experiences with which they would be allowed to personalize their own knowledge and hence the visualizations. It is important to note that this claim is grounded within the findings of the study which we have presented in this paper.

## References

- [1] Borner, K., Chen, C. and Boyack, K. (2003) 'Visualizing Knowledge Domains'. *Annual Review of Information Science and Technology*.
- [2] Chen, C (1999). Visualising semantic spaces and author co-citation networks in digital libraries' *Information Processing and Management*, **35**(3), 401-420.
- [3] Chen C. and Paul R. J (2001). Visualizing a knowledge domain's intellectual structure'. *IEEE Compute*, **34**(3), 65-71.
- [4] Chen, C (2004). Searching for intellectual turning points: Progressive Knowledge Domain Visualization' In proceedings of the *National Academy of Sciences of the United States of America* (PNAS), **101**(1), 5303-5310.
- [5] Fairchild, K.M., Poltrock, S.E. and Furnas, G.W.(1988) SemNet: Three-dimensional representations of large knowledge bases'. *Cognitive Science and Its Applications for Human-Computer Interaction*. Guindon, R (Eds), 201-233.
- [6] Mackinlay, J. D., Rao, R. and Card, S. K. (1995) An organic user interface for searching citation links'. In Proceedings of the SIGCHI conference on *Human factors in computing systems*, Denver, Colorado, United States, 67-73, ACM Press/Addison-Wesley Publishing Co, New York, NY, USA.
- [7] Shneiderman, B., Feldman, D., Rose, A. and Ferr Grau, X. (2000) Visualizing digital library search results with categorical and hierarchical axes', In proceedings of the fifth *ACM conference on Digital libraries*, San Antonio, Texas, United States, 57-66, ACM Press, New York, NY, USA.
- [8] Strauss, A. and Corbin, J. M. (1998) *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*, SAGE Publications.