

# Challenges of Evaluating the Information Visualisation Experience

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

Information Visualisation (InfoVis) is defined as an interactive visual representation of abstract data. We view the user's interaction with InfoVis tools as an experience which is made up of a set of highly demanding cognitive activities. These activities assist users in making sense and gaining knowledge of the represented domain. Usability studies that involve a task-based analysis and usability questionnaires are not enough to capture such an experience. This paper discusses the challenges involved when it comes to evaluating InfoVis tools by giving an overview of the activities involved in an InfoVis experience and demonstrating how they affect the visualisation process. The argument in this paper is based on our experiences in designing, building and evaluating an academic literature visualisation tool.

## Categories and Subject Descriptors

H5.m. [Information interfaces and presentation] (e.g., HCI)

## General Terms

Human Factors

## Keywords

Information Visualisation, User Experience, Evaluation.

## 1. INTRODUCTION

Information Visualisation (InfoVis) is the visual representation of abstract data of a specific domain on a computer screen. We consider user's interaction with InfoVis systems to be an experience, due to the activities that users are engaged with in addition to the knowledge being gained. This experience is made out of a combination of cognitive activities which are related to making sense and gaining knowledge of the visually represented domain, and a set of activities which are related to interacting with the interface. Standard HCI usability measures rely mainly on capturing the usability of the interface with no clear manner with which to capture domain related cognitive knowledge. Hence by merely relying on standard HCI usability measures, insight into an experience is fragmentary, since only the usability of the interface is captured. In this paper we argue that in order to evaluate InfoVis tools, we need to take into account both of these activities due to their interconnectivity. In other words,

we need to capture the user's visualisation experience as a whole. The argument we make is based on our experiences in designing and evaluating an academic literature visualisation tool, drawing on observations and analysis from a pilot study.

## 2. THE INFOVIS EXPERIENCE

Spence [9] describes InfoVis as a cognitive activity with which users are engaged with the potential of gaining an insight and an understanding of the represented data. This participation of the user in a cognitive activity leads not only to gaining domain related knowledge but is the substance of the user experience with the visualisation. In addition to this knowledge gaining activity, the user is also involved with activities that are essential to interacting with the represented tool. It is the combination of both that forms the user's visualisation experience. Hence when it comes to evaluating InfoVis tools it is essential to capture the experience as a whole. The user's visualisation experience, whether it is related to the gaining of domain knowledge or interacting with the interface, has a cognitive nature related to mental images of the domain and user's mental model of the interface. The interconnectivity and cognitive nature of the user's mental image and interaction model makes the user experience hard to measure.

## 3. THE COGNITIVE ACTIVITIES

The main user role in the InfoVis process is a cognitive one. The user interprets the visual representations of the data and builds mental images from which knowledge of the domain is gained. The visual representations are the only means with which domain related knowledge can be communicated to the user. Users need to make sense of these visual representations, hence engaging in cognitive activities. Some authors claim that InfoVis has moved out of the mind and onto the computer screen [10]. This is true in the physical sense, however, the activities involved from the user are still very much in the mind. In addition to users' engagement with these visual representations, users must also interact with the interface itself, in order to do so they rely on mental models that they develop of the system. This concept has been thoroughly covered in HCI literature and reflected in standard usability measures. However, when it comes to the building of domain related knowledge, there is a substantial lack in the literature. Hence, this forms the main challenge when it comes to the evaluation of InfoVis tools.

### 3.1. Mental Models: InfoVis and HCI

Both InfoVis literature and HCI literature refer to cognitive activities that users engage with in order to gain knowledge and interact with a system. This is reflected by the concept of building mental images of abstract domains or models of the interface. However, there are crucial differences between InfoVis and HCI literature when it comes to defining these concepts. In InfoVis literature, mental images represent the

images that people construct of a particular domain, whereas in HCI, it represents the models that people develop of a system [7]: from this point forward we will refer to this HCI model as the interaction model. We argue that mental images and interaction models cannot be separated since they both take part in the user's InfoVis experience, and hence must take part in the evaluation process of InfoVis tools. The visualisation activity is seen as that of building an internal interface [10] that cannot be printed or seen by anyone other than the user. It is through this act that users gain knowledge and insight of the represented domain.

### 3.2. InfoVis Mental Images is the Challenge

When it comes to mental images, in the context of InfoVis, there is no right or wrong: it is how people make sense of something they interact with. Since it is their belief, it is not susceptible to rigorous tests. The capturing and validating of mental models is a difficult task. In addition, it is a controversial one: Rogers et al [8] have proven that merely trying to make people talk about their mental model may in fact affect and change these models, proving its delicacy. In order for users to create mental images of the domain whilst interacting with an InfoVis system, they engage in a number of activities such as: interpretation of the visual cues, building associations, identifying similarities, etc. In addition, users rely on the models they build of the system's interface in order to interact with the InfoVis. Whilst in this interaction process they are engaged in the process of building domain related models. Interaction models and the building of mental images of the domain are interrelated and hence must take part in the evaluation process of InfoVis tools. Ideally, interaction models should not interfere with the building of domain related mental images; on the contrary, they should complement it. Users face challenges when interacting with InfoVis tools which rely on the fact that information cannot be seen: it is interpreted by users from the represented data.

## 4. EVALUATING INFOVIS TOOLS

In InfoVis the raw data itself is not the goal, it is the information it conveys. It is important to note that information and data are not equivalent. Information is derived from the data as Spence [9] indicates. By looking at the visually represented data, in other words browsing through the representation, interesting information is revealed. This exploration results in gaining higher levels of knowledge at the semantic level. However, before such knowledge can be gained users must interact with the interface at a syntactic level through a set of visual tasks, such as: identifying individual entities, categorizing entities, identifying clusters, etc. In order to evaluate InfoVis tools we should target knowledge at both the syntactic and semantic levels through a set of low and high level tasks. Usability in general is not a standard practice when it comes to designing and building InfoVis tools. Of the studies that do exist we categorize them according to the level of knowledge that they target, in other words, whether they target knowledge at the syntactic level or the semantic level.

### 4.1. Syntactic-Knowledge: Low-level tasks

Low-level tasks represent the tasks performed by users at the syntactic level of the InfoVis user experience. The low-level tasks evaluate whether or not the user understands the syntax of the visual language. Zhou and Feiner [11] identified a visual task taxonomy from which domain independent visual tasks were identified. Examples of these tasks are: identify, locate, rank, generalize, correlate, etc. Morse et al [6] used

these low-level tasks to devise specialized tasks which were used to evaluate visualisations at the syntactic level of knowledge.

### 4.2. Semantic-Knowledge: High-level tasks

The high-level tasks correspond to the tasks used to evaluate the visualisation at the semantic level. They are tightly coupled with the visualized domain, unlike the low-level tasks. Evaluating the visualisation using these tasks assists in determining whether the visualisation design corresponds to the requirements of the tool. However, unlike the low-level tasks, where the primary tasks are generated regardless of the represented domain, the high level tasks are generated from user requirements. Kobsa [5] argues that simply giving users low-level tasks, such as searching for a specific entity, or performing counting tasks makes it easier to identify the usability issues, since tasks such as these are rarely executed by users in a real InfoVis experience situation. When it comes to InfoVis systems there is more than one way of executing a task; as a result they emphasize that testing for high-level tasks is of great importance. We completely agree with this argument, as we discuss next. However, we strongly believe that both the low-level tasks and the high-level tasks are interconnected, since in order for users to execute the high level tasks they must go through a set of low-level tasks. Tasks at both of these levels represent activities that take part in the building of domain related mental images, they however do not target interaction related knowledge.

## 5. THE ACADEMIC LITERATURE VISUALISATION TOOL

The challenge of evaluating an information visualisation has been made concrete for us by our work in visualising academic literature [2]. This domain consists of the literature data within an academic context. It includes information such as: authors, papers, citations, journals, etc. Users of such information are mainly researchers in an academic field since it is important for them to keep track of the literature. In addition, they also need to create a complete and global understanding of the community. Literature data is complex due to its size and interrelations that appear between the entities, e.g.: citation trails where a paper cites another paper which in turn cites another, etc. Researchers would need to keep track of thousands of literature items ranging from the authors of the publications to the detailed ideas presented in each publication. The diversity and individualism of academic literature users formed one of the main challenges in the design and evaluation of the associated visualisation tool.

### 5.1. The Design

Prior to designing the literature visualisation tool we conducted a qualitative study [2] to capture users' experiences whilst interacting with academic literature. The qualitative study 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 academic literature. Hence, the goal is to give the user the freedom to explore and manipulate the visualisation through multiple interactive activities. In addition, users need to be given the ability to personalize their experiences. We developed a preliminary prototype of a literature domain [3] (Figure 1). 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 [4]. The data layout and interactive activities were based on users' literature experiences.

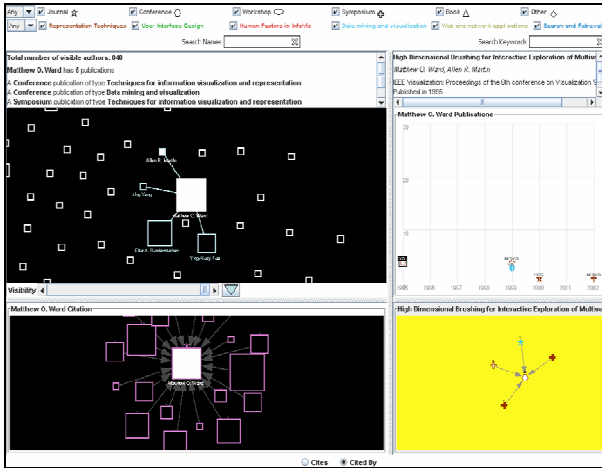


Figure 1. Prototype of the literature visualisation tool

## 5.2. The Evaluation

As discussed earlier, we believe that InfoVis tools must be evaluated at the syntactic level before they can be evaluated at the semantic level through a set of low-level tasks that target the visual syntax. The generated low-level tasks must fit within the context of the interface, since it is through interacting with the interface that any InfoVis knowledge can be learned. Therefore, we took a similar approach as Morse et al [6], where we devised tasks based on Zhou and Feiner [11] visual taxonomies.

Table 1 Difficulty level and its associated visual task

Difficulty	Visual Task
Primitive	Locate, identify, associate
Intermediate	Categorise/rank, compare/distinguish, reveal, cluster, Correlate
Complex	Generalise

Based on our system's design rationale we decided that the following low-level tasks best represented our needs: locate, identify, associate, categorise, rank, compare, distinguish, reveal, cluster, correlate and generalize. Unlike Morse et al [6], we ranked these tasks into three levels of difficulty (Table 1). The difficulty levels we identified were based on the primitiveness of the actions that the users had to engage with in order to accomplish the specific task, which in turn related to the complexity of the visual language syntax. The reason for doing so is that we are interested in identifying usability problems which are related with the visual syntax, in addition to the user's interaction model. For each of the actions we devised a set of tasks. These tasks were based on the results of the qualitative study we conducted [2].

### Primitive tasks

We identified the following actions as primitive:

Locate tasks: assist us in determining whether or not the user understands the visual data layout since it is essential that they are able to locate specific entities. For example: Locate [author\_name]'s 2001 paper, what are its keywords?

Identify tasks: assist us in determining whether the user is able to understand the visual encodings, such as shape and color, which are used to encode the data. For Example: Of the

papers that have been published by [author\_name] which has been cited the most?

Associate tasks: assists us in determining whether the user understands the relationships between the various visual entities. For example: Which research interest is associated with more authors?

### Intermediate Tasks

From these primitive actions more complex actions were identified based on Zhou and Feiner [11] visual taxonomies: categorize/rank, compare/distinguish, reveal, cluster and compare. We base our categorization on the fact that in order for the user to be able to accomplish any of these actions one or more of the primitive action must be executed. For example: The 1996 [paper\_title] by [author\_name] cites another paper by [author\_name], who does the latter paper cite? This is an intermediate revelation task since it is made out of various primitive actions, which are: *locate* and *identify*. In order for the user to accomplish this task the user must first *locate* the particular author, *identify* the specific paper and then reveal its citation information.

### Complex

Complex tasks lean more towards the high-level tasks since they are based on users' exploration of the visualisation and not on specifics. They are less controlling compared to the primitive or the intermediate tasks in terms of the actions that the user must execute, and the answers they provide. Complex tasks are related to the 'generalize' visual tasks since users reach generalized conclusions on the entities being explored. An example of a complex task: What is the relationship between the following authors [author\_name] and [author\_name]? In this task the users are given the freedom to explore the visualisation to identify the relationship. Comparing this task with an intermediate task such as: [author\_name] cites [author\_name], how many papers did they co-author? In this task the users are restricted in the way they would explore the visualisation since the grounds of the relationship are specified. However, with the complex tasks different users can reach different results in various ways. A few users identified a relationship between the authors that we had not previously come across despite extensive knowledge of the dataset.

## 5.3. Case Study: Test of Evaluation

In order to evaluate our evaluation approach, we conducted a pilot-study which took the form of a standard HCI task-based evaluation study which was based on the tasks devised. The tasks were given by order of difficulty: primitive, intermediate and complex. In total a set of 22 questions were given to the participants. During the course of the study the researcher was taking notes in addition to measuring the time participants needed to answer each question. Prior to beginning the study demographic information was gathered in addition to participants' knowledge and experiences with InfoVis tools in general and literature visualisation tools in particular. Users' interaction satisfaction was captured using the Questionnaire of User Interface Satisfaction (QUIS) [1]. There were seven participants in total. Six of them had at least two years experience in doing research. Two of these participants did not know what InfoVis was. One participant was an expert in the field of InfoVis. None of the participants had ever worked with literature visualisation tools. The questionnaire helped us identify some minor usability issues that needed improvement.

## 6. WHAT WE LEARNT

This study did not reflect anything that relates to the users' InfoVis experience. This was expected since the aim of the study was to capture the users' syntactic knowledge and interaction model. However, a lot was learnt in relation to the evaluation of InfoVis tools.

### 6.1. Low-level Tasks: Usability not Experience

A controlled task-based usability study assisted us in evaluating whether or not the system's interaction model fitted the users' interaction model which, as we discussed earlier, is an essential part in evaluating InfoVis tools. However, it said nothing about the visualisation or the experience as a whole. In fact it seemed to hinder such an experience. The study revealed that following a standard task-based evaluation method restricted users' experiences. For example, one of the participants commented: "I liked the system but I was trying to complete tasks given – it might mean more if I used it for research and saw interesting relations in papers that was meaningful to me".

We merely got a glimpse of the users' experiences with the visualisation tool through the comments that the users gave and the observations the researcher captured during the study. We argue that relying on quantitative measures proved to be insufficient. We strongly believe that relying on higher-level tasks and the combination of qualitative and quantitative analysis will assist us in capturing users' InfoVis experiences more effectively.

### 6.2. Efficiency: Physical Activities rather than Time

Users were timed during the pilot-study as they answered each of the questions. After analyzing this data there was not a straight correlation between how well the users did and the time it took them to do it. In fact, it seemed that the more time they spent on each of the tasks the better they did. However, due to the number of participants we cannot unequivocally assert the claim. But we can claim that, from our observation of the users' performances and experiences during the study, there was a direct relation between the number of physical activities that they engaged with and their overall satisfaction, which reemphasize our argument that the execution of the interaction model should complement and not interfere with the creation of the domain related mental images. We observed that the more they had to engage with physical activities as: constantly clicking, or zooming in and zooming out to accomplish something specific the more they were frustrated, as observed by the researcher through: the comments that they gave, the facial expressions or gestures they performed. They seemed to want to have the information they needed, that related to a specific task, with the least number of physical activities. We hypothesize that the amount of physical activities users engage with whilst accomplishing a specific task might be a better measure of efficiency rather than relying on the time it takes users to accomplish a task.

### 6.3. Experience is the Essence

Interestingly, all complex questions were answered correctly by the participants. This might be due to the fact that these tasks come last in the list of questions given to participants. In addition, it might be due to the fact that these complex tasks, as expressed earlier, are less restrictive compared to the primitive and intermediate tasks. As a result they give users the freedom to explore the InfoVis interface, hence allowing for a better experience. This is not an assertion but a

hypothesis that needs to be further investigated. One of the participants commented: "It got easier to use with more practice, also became more adventurous...". From here we identified interesting questions: What makes it "more adventurous", what makes it a better experience? This is what we would like to capture and understand.

## 7. CONCLUSION

Where does HCI lie when it comes to evaluating InfoVis tools? It assisted us in successfully identifying usability issues related to the user's interaction models. However, the InfoVis experience is not just made out of the interaction models but it also relies on the building of domain related mental images. Mental images are built as the user interacts with the visualisation at the syntactic and semantic levels of knowledge. From our pilot-study we argue that in order to evaluate the InfoVis we need to capture the experience as a whole, which remains as a challenge to the field of HCI. In addition, we also demonstrated that efficiency in the InfoVis tool cannot be captured using time and suggested the use of physical activities instead.

## 8. REFERENCES

- [1] Chin, J., Diehl, V., Norman, K. (1988) 'Development of an instrument measuring user satisfaction of the human-computer interface', Proc of SIGCHI conference on Human factors in computing systems, 213-218.
- [2] Faisal, S., Cairns, P., Blandford, A., (2006) 'Developing User Requirements for Visualisations of Literature Knowledge Domains,' 10<sup>th</sup> International Conference on Information Visualisation (IV'06), 264-269.
- [3] Faisal, S., Cairns, P., Blandford, A., (2007) 'Building for Users not for Experts: Designing a Visualisation of the literature domain,' to be presented at 11<sup>th</sup> International Conference on Information Visualisation (IV'07).
- [4] Ke, W., Borner, K. and Viswanath, L (2004), InfoVis04 Contest MS-Access Database, Indiana University, School of Library and Information Science and School of Informatics, <<http://ella.slis.indiana.edu/~lviswana/iv04-contest.mdb>>
- [5] Kobsa, A. (2001) 'An Empirical Comparison of Three Commercial Information Visualisation Systems'. IEEE Symposium on Information Visualisation 2001, 123-130.
- [6] Morse, E., Lewis, M. and Olsen, K. A. (2000) 'Evaluating Visualisations: Using a Taxonomic Guide', *International Journal of Human-Computer Studies*, 53(5), 637-662.
- [7] Norman, D. (1988) *Psychology of Everyday Things*. Basic Books
- [8] Rogers, Y., Rutherford, A., and Bibby, P. (1992). *Models In the Mind - Theory, Perspective, and Application*. London: Academic Press.
- [9] Spence, R. (2001) *Information Visualisation*, first edition ACM Press Books, Edinburgh Gate, Harlow, Essex, UK.
- [10] Ware, C. (2000). *Information Visualisation: Perception for Design*: Morgan Kaufmann.
- [11] Zhou, M., Feiner, S. (1998) 'Visual task characterization for automated visual discourse synthesis', Proceedings of the SIGCHI conference on Human factors in computing systems, 392-399.