VISUALIZING FOR EXPLORATION AND DISCOVERY

Tanja Merčun¹
Maja Žumer²

Abstract
In the last few years, libraries have been faced with a harsh reality in which they have been losing their users and their position as a primary information source. In the attempt to improve their services, we have witnessed a growing number of modernized library information systems, tools, and features and while libraries (or at least some of them) have made a significant step forward in this area, there is still a need for improved exploratory and discovery tools. The paper first looks into the concepts of discovery, serendipity, and exploration and how they are supported in the new generation library systems. In the second part, possible role and application of visualization for exploration is discussed and examples given from a number of existing online visual tools and services. The last part then reflects on potential uses of visualization in library information systems and gives some guidelines on designing visualization for discovery. The authors conclude with the thought that libraries should couple visualization with textual data and implement simple and small-scale visualization that would give users some more insight into the library data and at the same time invite them to explore the collections.

Keywords: exploration, discovery, library information systems, information visualization

1 INTRODUCTION
Today, users have high expectations from all services provided to them and therefore have no problem turning elsewhere when a service does not meet their strict requirements. This behaviour is not typical only of the young, but rather of all population that uses the web on daily basis. Faced with this new reality, where users have chosen the rich online environment over traditional library services and systems, libraries have been forced to reflect on their current services. With the demanding and unforgiving google generation users and the rapidly developing technology, libraries have realized they needed to improve their online presence and services and make better use of their data. Libraries do hold and produce a vast number of valuable data that could be better exploited and used to libraries’ advantage over other information providers. We know that going beyond simple known-item search systems is a must and developing more effective and innovative tools for exploration is one of the areas where libraries could put their specific knowledge and data into good use. Faceted navigation, suggested titles, and tag clouds have proven to be one possible way to enhance users’ navigation and exploration, but with the growing numbers of library materials and users, who demand easy access and high performance systems, there is still a need for further development. We believe that with the people’s growing preferences for visual information and the potentials of information visualization for exploration, libraries should look into ways to visualize at least some of the data in their information systems.

2 BEYOND KNOWN-ITEM SEARCH: THE NEED FOR EXPLORATION AND DISCOVERY
In current library information systems, users do not get full benefit from library data. The reason for that is twofold: while there is no doubt that more could be done to improve the quality of the data, our current systems in many cases do not really use the full potential of the data that we do have. In order to improve the systems’ ability to support exploration and discovery, changes need to be made on
both levels. This paper, however, will mainly deal with the latter, focusing on how presentation of results in user interface design could enhance exploration and discovery.

2.1 DISCOVERY & SERENDIPITY

Exploratory information seeking not only supports users’ vague information needs, but can also lead to unexpected, accidental discovery of something valuable or what is often called serendipity or serendipitous discovery. Although research on serendipity in the information seeking process has been rare and serendipity has not been given a visible place in the information seeking models, it has been recognized to bear important potential in the exploratory information seeking process (Foster & Ford, 2003). The definitions of serendipity show that people often encounter it when seeking for something completely different, thus finding unexpected, needed, or just interesting information that can result in an unexpected change in the direction of their search (McBirnie, 2009).

André et al. (2009) define serendipity as:

1) the finding of unexpected information (relevant to the goal or not) while engaged in any information activity
2) the making of an intellectual leap of understanding with that information to arrive at an insight.

In their paper André et al. (2009) argue that while the second half of definition depends largely on an individual’s previous knowledge and ability to use information and make the connection, we should design our systems to aid the first part of the serendipity process: the discovery of new pieces of information that would most likely encourage the creation of unexpected connections.

2.2 EXPLORATORY INFORMATION SEEKING

Exploratory information seeking can be defined as a class of search activities that move beyond fact or known-item retrieval towards fostering learning and investigation, generally using a combination of querying and browsing strategies (White et al. 2006a; 2007). Also Wilson et al. (2007) mark investigation and learning as the two most important actions in exploratory information seeking, both requiring large user involvement and a range of activities such as comparison, aggregation, and evaluation.

Designing exploratory search systems, we therefore need to think about ways to support this interactive and dynamic process between the user, the system, and the information sources. As Schraefel (2009) well puts it, our tools must go even further than “people who read a also read b” and tell searchers what they do not know but should.

Considering the question of how user interfaces should support exploratory information seeking, Marchionini (2006) suggests tools for:

a) query development and improvement,
b) navigation,
c) organization,
d) visualization.

There are also others that suggest some or all of these features. Authors such as Kules (2006) and White et al. (2007) promote organization and navigation as two of the basic elements for exploratory information seeking. As they explain it, meaningful groups or categories have a good potential for supporting user exploration, understanding, reflection, and discovery by providing a semantic roadmap to knowledge creation and improved learning (Kules 2006; Kwasnik, 1999; Kules & Shneiderman, 2008; Koshman et al., 2006). White et al. (2006b; 2007) further argue that navigation is a crucial part in supporting exploratory search as users benefit from interfaces that offer browsing, dynamic queries, guided navigation, hypertext, and visual representation of the space. Also Shneiderman (2008), Kules (2006), and Schraefel (2009) see potential in integrating interactive information visualization to support insight, facilitate rapid refinements of queries and allow users to research questions that are not very easily expressed in keyword search.

Anick (2008) and Schraefel (2009) add to the basic list some additional supporting tools such as: longer snippets (for supporting the learning phase in exploratory search), recommendations, tools for marking and remembering useful documents, and tools for collaborative knowledge building, all of
which are not really exploratory tools in themselves, but can support the investigating and learning process in exploration.

3 SUPPORTING EXPLORATION IN NEW GEN LIBRARY SYSTEMS

In new generation library systems we have witnessed some interesting features that have the ability to support exploration and discovery. Unfortunately their creation cannot really be contributed to libraries themselves as most of the tools first appeared on the web and have only later been recognized and adopted by libraries.

One of the most noted exploratory tools in the last few years has been the faceted search (e.g. Wilson & Schraefel, 2008; Kules, 2006), which has, according to Lemieux (2009) and Medynskiy (2009), become a de facto standard and so ubiquitous that users are coming to expect it. Ben-Yitzhak et al. (2008) describe it as an interaction paradigm for discovery and mining, where the user is able to start search, refine a search query, or navigate through multiple independent categories that describe the data by drill-down (refinement) or roll-up (generalization) operations. But while faceted navigation has most often been noted as a technique that supports exploratory information, Zhuge & He (2009) argue that it is poor in schema flexibility and lacks rich semantics needed to fulfil the exploration task, such as advanced operations for exposing relations among resource sets. We believe that information visualization could fill this void.

Besides faceted navigation, there are also other features we can describe as exploratory, such as recommendations and featured items lists, query suggestions, and social navigation. University of Huddersfield library catalogue has even developed a visual shelf that enables users the kind of browsing they would experience in a physical library.

3.1 EXPLORATION AND DISCOVERY USING VISUALIZATION?

As we have reported in previous chapters, visualization is believed to be a promising tool for exploring data in order to gain understanding of the data and the phenomena behind them (Purchase et al., 2008; André et al., 2009; Beale, 2006). The argument for the use of visualization is therefore that it presents an interactive mechanism for browsing, exploring, and analysing which increases people’s ability to perform these activities, thus helping them overcome large information spaces, build new knowledge, as well as discover and understand relationships and the information space.

The idea of information visualization has been around for more than two decades, but even though visual presentations are supposed to enable people to process the data quicker and better by diminishing the cognitive load, studies so far have not been able to really prove its superiority over linear information presentation. What is also concerning is that already established information visualization systems such as Grokker and Kartoo closed down in the last couple of years.

Admitting that visualization attempts in library systems and elsewhere have not yet proven very successful, one could conclude that visualization may not be the solution we are looking for. However, the discouraging results may not necessarily lie in the visualization concept itself, but in the way visualization was designed and implemented. As Kules (2006) wrote, designers have too often experimented with all the possibilities of what could be done instead of creating what should actually be done. While it is difficult to make such claims without a proper in-depth study of existing experiments, we must realize that poor design and usability have very likely attributed to the current status of visualization.

4 EXAMPLES OF CURRENT VISUALIZATIONS ON THE WEB

To get a better understanding of the current use of information visualization in search systems, we will look at some examples currently available on the web (figures can be found in Appendix).

**EYEPLORER**

Eyeplorer (Figure 1) searches Wikipedia and MEDLINE/PubMed database, processes the retrieved sites, pulls out the main concepts and organises them in categories. Holding a mouse over a node displays an explanation on how the concept relates to our query and clicking on a term shows relations with other concepts. These relations are not only indicated with lines as it often happens in visualizations, but can also give a textual explanation as clicking on a link brings
forward a description of the connection. The system is quite novel and managed solely by computers, so the results and descriptions vary in their quality and accuracy. Nevertheless, we can say that Eyeplorer presents a good idea of what information visualization system should offer: it gives clear overviews of the most important topics, shows and even explains relations and enables users to interact with the visualization.

LIVEPLASMA & TUNEGLUE

Both, Liveplasma and Tuneglue show clusters of related entities. Similar to Eyeplorer, which advertises its services with a slogan “Explore relationships”, the two tools also describe themselves as exploratory in nature: Tuneglue as “relationship explorer” and Liveplasma as “discovery engine”. Tuneglue (Figure 2) focuses only on musical artists and displays entities that are related to the chosen artist/band. It does not, however, offer the user much interactivity with the results, take advantage of visual properties (such as colour, size, proximity) or give information on the nature of relations between entities. We might assume that if we liked one artist, we would also like related artists, but beyond that relationships are not made clear to the user. Whether they were made on the basis of music genre, popularity, time period, or some other attribute therefore remains unanswered.

Liveplasma on the other hand, offers music as well as movie exploration. In displays for music (Figure 3), users are presented with the artist/band nodes of different sizes according to the popularity of the artist but again no explanation is given on the nature of relationships or the meaning of node’s colour. For movie search (Figure 4), the situation is somewhat better as the legend explains which colour belongs to which movie genre and how the line colour is used to differentiate between the person’s roles in a movie: violet line for movie director and yellow line for actors. When searching for movies, the retrieved visual display shows related movies and movie directors, and when searching for a director/actor it presents the person’s opus. Both Liveplasma and Tuneglue enable interaction with the visualization, Tuneglue only to the extent that the user can expand a chosen entity while Lieplasma enables expansion as well as formulation of a new query by clicking on the chosen entity.

MUSICOVERY

Another tool for music discovery is Musicover (Figure 5). The most interesting feature it offers is certainly the visual browsing bar where it is possible to define either the mood or dance tempo within a kind of coordinate system. Depending on the chosen value, the system then displays results, colouring them according to the music genre. The visual display of results itself is not a visualization that would produce any additional information, interaction options for the user, meaningful relations, or use other visualization parameters to communicate information, so the value of the system really lies in its ability to browse using the visual mood/tempo feature.

TOUCHGRAPH

Visualizations for books are few and hard to find on the web. Touchgraph is a visualization tool that provides online visualizations for movies, music, and books based on Amazon data. Figure 6 presents an example of results display that has been expanded to more similar items by clicking on some of the nodes. The size of the cover reflects the sales rank and colour is used to show clusters where an item is a recommended purchase for another item. Using the visualization itself, users can only expand the search results, while they need to use bars on the left in order to refine the search. On the left there is also a detailed description of the chosen result, but as with other visualizations, it is still difficult to guess for a user how some item relate to each other or what exactly are the differences between items with the same title based solely on the visualization.

The presented tools deal with data similar to our library information systems. While their final aim is, as they all claim, exploration and discovery, there are some differences in their approaches. Some systems are quite simple and limited in the amount of presented data (for example Tuneglue only shows the names of related artists) while others provide somewhat more information either on the
chosen nodes or the relationships between them. Beside the content level they also differ in their execution. In opposite to all others who visualized result lists, Musicoverry for example decided on using information visualization as a dynamic query option. Further looking at the options for interaction and modification, Touchgraph and Eyeplorer used the most options that enable users to not only refine the contents, but also to control the display itself. In Touchgraph users have the possibility to determine the size and colour of clusters and spacing between the nodes and in Eyeplorer they can drag and drop interesting explanation from the visualization into their notebook, drag a node into a query field and thus form a new refined search, they can also choose the number of concepts they wish shown or just the cluster that they want to take a closer look at (for example only Science or People cluster). In all systems except Eyeplorer we have missed some sort of explanations on both the content and the way visualization was constructed as the lack of that information gave us the feeling that we cannot really take the full advantage of the system.

While some visualizations were better planned and executed than others, they all showed us several problems we can encounter in design of visualizations. One is the lack of information that would enable users to understand the visualization (such as explanations of relations, used colours…) and the other is the lack of textual data that would give more information about the domain and thus bring users new knowledge and support the exploratory process. Many of the presented examples exploited only a fraction of visualization possibilities, which goes against Tufte’s (1990) fundamental rules for visual displays, which says that every pixel on the screen should be information bearing. The presented visualizations also offered little interaction and customization options, thus limiting the amount of information that visualization could communicate to the user.

A quick look into a few visualizations found on the web showed that making visualizations work is a difficult task that requires a well-planned information architecture and interaction design as well as appropriate visual presentation. We must also consider the quality of our underlying data (which we know is likely to be a problem in many library databases) as, similarly to other groupings and automated procedures, small mistakes or even inconsistencies are quickly revealed in visualized displays and may misguide users or take away some of the usefulness of visualization.

5 POSSIBLE APPLICATIONS OF VISUALIZATION FOR EXPLORATION IN LIBRARY SYSTEMS

When thinking about visualization in library systems, we should consider two aspects:

a) what functions/features could benefit from visualization and
b) how should visualization be designed in order to be enable easy and intuitive use.

5.1 FUNCTIONS THAT COULD BE VISUALIZED

Looking at library catalogues, we have seen that visual information is gaining importance with simple visualizations such as tag clouds, time sliders for refining search, featured cover displays of new acquisitions, recently borrowed items, book shelf display etc., and even a sort of vocabularies (“word cloud” in Aquabrowser). Considering other visualization possibilities, there are various options as to how visualisations could assist and encourage user’s exploration and discovery in library information systems. The key with visualizations is that we use already existing relationships or relationships that can be derived from our data to improve methods for displaying, organising, and browsing information (Kipp, 2008). So far we have not really put the mass of data already available in our databases to a good use, which is why visualization seems a good way to make our data work harder (if we borrow Lorcan Dempsey’s phrase). That is why we are currently researching the possible use of visualization for the display of facetized records (see Merčun & Žumer, 2009) that could present a network of relations connected to the chosen entity (for example expressions and manifestations of a work, works deriving from or preceding a work, related authors and topics etc.).

There are also other possibilities that, to some extent, depend on the type of system, the content it holds and users it serves. The needs of an academic or governmental library system are, in many cases, completely different from systems such as public library catalogues or digital libraries. Academic and research information systems could, for example, benefit from bibliometric visualizations that have already been quite extensively tested in various prototypes. On a more general note, we could visualize data such as circulation, personalized suggestions based on user’s search history and borrowed items, or thesaurus display and subject categories. In 2.0 participatory
systems, visualizations could also be applied to social features such as user-created lists and user profiles, making a network that would help people find other potentially interesting books or users. In case we had full text, visualization of content would be useful; it would also be really exciting to implement something similar to Musicover’s mood/tempo navigation with different values according to media and type (for fiction works maybe attributes such as sad-happy ending, easy-difficult read), but that would probably require some additional metadata descriptions and manual positioning. Similarly to the examples we presented in the previous chapter, library systems could also enable users to explore similar artists, writers or movies via a visualized presentation. We could even introduce geovisualization to our systems and allow users to explore literature or music via maps.

Generally looking at the features and functions, visualization can aid the exploratory process in various levels of information seeking. By its function, we could say that visualization can take up different roles. It can be the central point of the system, presenting overviews of the collection and search results, allowing us to browse and explore records by interacting with the visualization or it can have an assistive or auxiliary function that visualizes other potentially interesting data or helps with queries or browsing.

5.2 HOW TO DESIGN VISUALIZATION?

Once we have decided on what functions or features could be visualized, we need to think about how to plan the system to reach its full potential. Designing a useful visualized display, we must consider three perspectives:
- textual data,
- visual appearance, and
- interactivity of the interface.

While the central point of visualization is the graphical presentation of data, planning the role of textual data is just as important. Kules (2006) even argues that it is the general lack of textual data in visualizations that has contributed to the disappointing results of information visualization. There are several questions to contemplate, such as: how to integrate textual data into what is essentially a graphical presentation of a large amount of data, what labels to use, and how much and what textual data is needed to support visualization. Only with some textual explanations will users gain more understanding of the visualization and knowledge of the explored topic. This may not be as necessary in visualization systems for experts on very specialized areas (where information visualization has really been most used so far), but when we are moving into more general environment such as libraries, we believe a combination of text and visual presentation is required.

For the visual appearance we choose the visualization technique, the size and granularity of visualization, and decide on parameters such as colours, node size, link strength and elasticity. The nature of underlying data (hierarchical, network, linear …) and the purpose of our visualization (do we want to expose links or do we want to emphasize the main trends etc.) help us determine which technique would best present the values of the dataset. The size and granularity of visualization are also important as displays can soon become illegible with all the visual clutter and make it difficult for users to perceive and comprehend all the elements on screen. Whether we display a general overview or only a smaller fraction of data, there should be only the amount of information that still stimulate human cognitive abilities. As for the colours, sizes, and proximity, they are the elements that can be used to communicate some of the information that should otherwise be explained textually.

Some implementations stop at the graphic presentation of data, a picture giving the user a general overview but missing, in our opinion, the crucial element that gives information visualization such potential: the interactivity - the ability for the user to interact with the display and learn through it. Also Ahlberg and Schneiderman (1994) write that the key principle of visual information seeking applications is to support browsing which enables rapid filtering to reduce result sets, progressive refinement of search parameters, continuous reformulation of goals, and visual scanning to identify results. The real power of visualization is therefore not in the passive displays, but in dynamic and flexible interfaces that enable users to indicate what is interesting and what is not and provide users with data that are needed for identification and explanations of the data that are understandable. Planning the visualization, we should therefore map and predict how the session can evolve with the human interaction and what customization options should be available to the user.
6 CONCLUSION

Visualizing information could be the next step in mining the data library holds either on the collection itself or the use of that collection. The biggest mistake with information visualization in the past has probably been the fact that visualizations have been too complicated to intuitively understand and very often too extensive to enable effective manipulation and exploration. What we would suggest at this point is that libraries try with small scale visualizations coupled with textual data and use them as supporting tools that invite users to further explore the collection and at the same time give them a better insight into library data. Much of the success lies in the hands of designers and librarians that need to make sure visualizations are well planned, attractive, useful, and usable. People are becoming increasingly visual beings that have come to expect interactive and discovery tools. If libraries want to keep up with other information sources on the web and offer their users the best possible service, they need to explore new possibilities and design services that will lead their user interfaces onto the next level.

BIBLIOGRAPHY


Figure 1: Visuализed display of results in Eyeplore groups results, indicates the node’s importance with size and provides also explanations of relationships.

Figure 2: Tuneglue display of related artists does not use the full potential of visualization options.
Figure 3: Liveplasma network display for related bands/artists uses proximity and node size parameters.

Figure 4: Liveplasma display for movie search uses colour to differentiate between relationships (director vs actor) and movie genres.
Figure 5: The value of Musicover lies in the visual touchpad for defining the desired mood or dance tempo, while the display of results serves a more aesthetic than informative purpose visualization should provide.

Figure 6: Touchgraph presents grouped networks of related books which the user can expand on or refine. Additional data on the chosen node are presented on the left handset, but the user may miss explanations of the established relations.