

Data Exploration for Bisociative Knowledge Discovery: A Brief Overview of Tools and Evaluation Methods

Tatiana Gossen, Marcus Nitsche, Stefan Haun, and Andreas Nürnberger

Data and Knowledge Engineering Group,
Faculty of Computer Science, Otto-von-Guericke-University, Germany
<http://www.findke.ovgu.de>

Abstract. In this chapter we explain the definition of the term (*data*) *exploration*. We refine this definition in the context of browsing, navigating and searching. We provide a definition of *bisociative exploration* and derive requirements on user interfaces, which are designed to support bisociative knowledge discovery. We discuss how to support subtasks of bisociative data exploration with appropriate user interface elements. We also present a set of exploratory tools, which are currently available or in development. Finally, we discuss the problem of usability evaluation in the context of exploratory search. Two main issues - complexity and comparability - are explained and possible solutions proposed.

Keywords: exploration, exploratory search, tools, usability evaluation.

1 Introduction

A lot of data in different domains, e.g. biology, astronomy, geography, and other sciences were gathered and became available during the last decades. Much useful and interesting knowledge is hidden in these data sets. Therefore, experts in different knowledge domains explore the data in order to make new discoveries and, thus, data exploration becomes one of the standard user tasks.

Unfortunately, the well-known phrase of the futurist John Naisbitt “We are drowning in information but starved for knowledge” [26] is still relevant. One man alone is not able to examine even small parts of the available data. Knowledge discovery tools are a way out and show a promising direction to support users by data exploration. Finding implicit links between given data (sets) from different domains is an even more challenging task. This is what bisociative knowledge discovery tools are supposed to support. This chapter addresses the issue of data exploration for bisociative knowledge discovery. Further details on bisociative knowledge discovery in general can be found in [6].

The structure of this work is as follows. In Sect. 2 we explain the meaning of (data) exploration, which, although often used, has no formal definition. Then, we discuss how to support bisociative data exploration through user interface

elements and give a short overview of tools for data discovery and exploration in Sect. 3. In Sect. 4 we discuss the problem of evaluating the performance and usability of knowledge discovery tools and propose a possible solution. The chapter ends with a conclusion and a discussion of future work in Sect. 5.

2 Bisociative Data Exploration

Many authors use the term *exploration* without giving a well-formed explanation. Since this term has many meanings and just some of them are related to computer science, it is important to clarify what exploration in the context of this chapter refers to. Therefore, we explain the meaning of the term *exploration* in the context of information retrieval, human computer interaction, visual data analysis and bisociative knowledge discovery. We also elaborate on the implications for user interface design.

2.1 Different Meanings of Exploration

A general definition of *exploring* is found in The American Heritage Dictionary of the English Language. Here, exploring is defined as “to search into or travel in for the purpose of discovery” [3].

In his book “Interaction Design” Michael Herzeg provides a more specific explanation of the term *exploration* in the context of human-computer interaction (translated from German into English from [19], p. 81):

“Exploration is a method for searching in new, unknown information spaces. It is similar to browsing with the important difference that exploration is defined by sounding the user’s interest in the information space. Further exploration is connected to the user’s wish to get to know almost the entire structure and the content of the information space. After a successful exploration of the information space users mainly navigate in it. The user builds up an overview map while exploring. The entry point for exploratory search is [...] in most cases given by accident.”

Contrary to Herzeg’s definition we see exploration not as a specific method of information access. More likely we see it as an enhancement of other methods: The basic information access methods *browsing*, *navigating* and *searching* can be enhanced by aspects of exploration. Hence, there exist *exploratory browsing*, *exploratory navigation* and *exploratory search*. Furthermore, the information space does not necessarily need to be *unknown*. For example, in exploratory navigation users are already familiar with certain navigation points and are able to make more sophisticated decisions where to go next.

Searching is often an integral part of exploration. Searching means “to look through (a place, records, etc.) thoroughly in order to find someone or something” [2]. Many users on computer systems are engaged in searching. Search systems usually provide an input box for entering search keywords to describe the user’s information need. Gary Marchionini [24] calls this *lookup* which is a

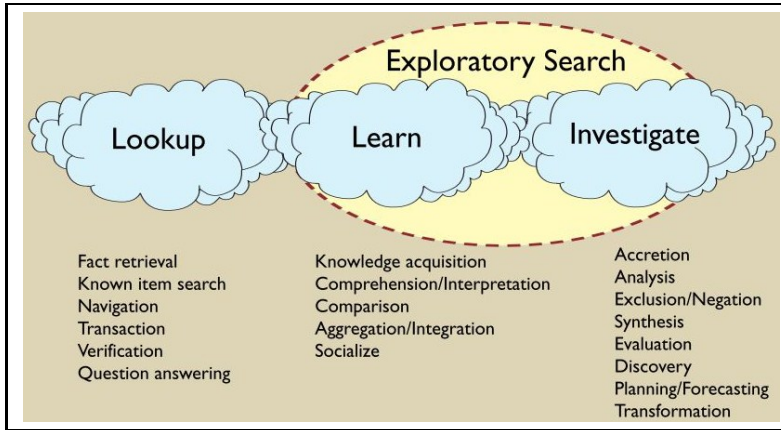


Fig. 1. Categorization of information access methods according to Marchionini [24]

summarization for procedures like fact retrieval, known item search, navigation, transaction, verification and question answering (see Figure 1).

Furthermore, Marchionini distinguishes between *lookup activity* and *exploratory search*. Exploratory search is, according to Marchionini, divided into *learning* and *investigating*. The first describes methods like knowledge acquisition, comprehension and interpretation, comparison, aggregation, integration and socializing. Investigative tasks contain accretion, analysis, exclusion and negation, synthesis, evaluation, discovery, planning, forecasting and transformation. Therefore the general task of exploration can be defined as exploring data sets in order to filter or extract relevant information from it, to (re)define the user's information need and to find associations between displayed information items.

Exploratory data analysis (EDA) is an approach to analyze data in order to formulate hypotheses worth testing and to complement programs of conventional testing for testing hypotheses [22]. We distinguish between this definition based on statistics and definitions found in human-computer interaction. EDA is a term named by John Tukey [38], which is strongly related to suggesting hypotheses about causes of observed phenomena in data and therefore has a specific use case in data analysis, while exploratory search is more about supporting the task in general.

Furthermore, there is a concept of *visual exploration*. Kreuseler [21] describes exploration as an undirected search for interesting features in a data set. According to John Tuckey [37], users engaged in exploratory data analysis are searching interactively and usually undirected for structures and trends in the data. Their goal is to find implicit but potentially useful information and they do not have an initial hypothesis about the data.

A definition for *visual exploration* is given by Tominski [35]:

“The aim pursued with visual exploration is to give an overview of the data and to allow users to interactively browse through different portions

of the data. In this scenario users have no or only vague hypotheses about the data; their aim is to find some. In this sense, visual exploration can be understood as an undirected search for relevant information within the data. To support users in the search process, a high degree of interactivity must be a key feature of visual exploration techniques.”

To sum up, we define *exploration* in the context of *searching* as follows: *Exploratory search is a highly dynamic process of a user to interact with an information space in order to satisfy an information need that requires learning about structure and/or content of the information space.*

Thereby the process expands a simple lookup by using techniques of exploration. Furthermore, users usually look at a (sub) set of information through a specific view angle, which might change during the investigation process. Therefore we call it *highly dynamic*. This personal and dynamic view is more generally known as the context of use that allows changes of perspective in order to (re)formulate or to refine an initial query. User’s overall goals are to learn, to investigate, to understand, or to conceptualize (about) their initial information need by building up a personal mental map or model. Thereby the acts of explorative searching, browsing and navigation are often more important than the actual find, i.e. success in this context does not necessarily mean to find a certain piece of information. This makes evaluation of exploration tasks rather difficult, as we discuss later.

2.2 Definition of Bisociative Exploration

In the context of this book we are talking about *bisociative knowledge discovery*. That means that people are engaged in the *creative* discovery of previously unknown information, in particular relationships that were before overlooked in-between different data sets. To find those *bisociations* users explore the data sets in a creative way: “Creative information exploration refers to a new information exploration paradigm that aims to facilitate the generation of creative insight or solutions.” *Bisociative information exploration* is an approach where elements from two or more “incompatible concepts” or information spaces are combined to generate creative solutions and insight [10]. Thus, we can define bisociative data exploration as follows:

Exploration is *bisociative*, if the data set consists of two or more habitually incompatible domains and the user is presented unusual, but interesting domain-crossing connections with the aim of finding relevant and new relationships between those domains.

2.3 Implications for User Interface Design

During exploration, the user is interactively exploring a given data set. His task is to build up a mental model - an overview map - in order to get a structured view on unknown data. Furthermore, the user needs to be able to follow connected

items in order to understand the context of the explored items. Another task, a user might try to fulfil, is to follow the boundaries of information he or she is already aware of to gain new insights and to establish further relations between the information items that are already known.

Since association is an important capacity in human communication, visual information addresses patterns of understanding [9]. Furthermore, exploration can be seen as a creative act that requires the user to recognize important parts of the presented information and build connections to the facts he or she already knows while discarding irrelevant knowledge or postpone it for later reference in a different context. This process can be easily disturbed by external influences such as distractions from the environments or inconvenient means of navigation through the information space.

As the user can only keep the most important seven chunks of information in his working memory [25] anything requiring one's immediate attention leads to a loss of just gained insight, which most probably has not been retained yet. Constantly losing ideas due to interruptions will make the user anxious and less productive. Not only one will lose information, but the higher probability of user's work being in vain will lower the motivation. Users may settle with worse results than the one initially intended to find. Supporting exploration in an ergonomic way means taking into account a lot of psychological and physical properties of human beings. Overall the requirements for bisociative knowledge discovery tools are:

1. Supporting *dynamic* exploration within information spaces.
2. Supporting users in deriving connections between different *domains*.
3. Supporting the human *creativity* processes.
4. It should incorporate *online mining tools* that are capable to return the results to the user interface in real-time, see e.g. [15].

Here we use the domain and creativity definitions from [10]. We believe that a tool that only visualize the whole data set, e.g. gives a large graph view, would not appropriately support bisociative knowledge discovery. The mentioned implications should be reflected in appropriate user interfaces. In the next section we give a brief overview of possible widgets for data exploration and discuss their applicability for bisociative knowledge discovery. We also present a short state of the art on existing exploration tools.

3 Supporting Bisociative Data Exploration

In this section, we first elaborate on the interface elements (widgets) which could be used to support data exploration. We analyse their applicability for bisociative data exploration. The second part of the section contains the state of the art on tools for data exploration.

There are several ways of providing users with appropriate user interface elements in order to support subtasks of exploration like *getting an overview*. Most common solutions are *lists*, *item sets* and *graphs*.

1. Lists are good to present accurate rankings, but due to their structure they only support sequential item observation. It is not possible to discover items in parallel or even generate an overview. Furthermore, relations between listed items are hard to identify without prior knowledge. Therefore lists do not provide good means for exploration tasks.
2. Item sets, often represented as Venn diagrams, provide a good overview about a data set. By using different colors and sizes, relations and groups can be easily recognized by users. Unfortunately, item sets do not provide users with an understanding where certain connections between single item sets exist.
3. Graphs consist of nodes and edges. While nodes often symbolize data entries or information items, edges provide users with understanding which connections / relations do exist between single items. Also groups of items can be identified by using different colors or sizes of nodes or by identifying separated sub graphs. One big disadvantage in this visualization method might be that users find it difficult to find an entry point to start from to explore the data set. Navigation might be also harder than in lists because there are multiple directions to go to. Especially for unknown data sets users might have difficulties to decide which path they like to follow. On the other hand exactly these characteristics support data sets analysis in an unconventional way and offer the possibility to find something new and unexpected, to find new insights in a certain topic.

Lists, item sets and graphs can be utilized to visualize the whole data set. They may also be incorporated within a tool for dynamic exploration of huge data, e.g. users can explore different levels of a collection which is visualized as a hierarchical list. But only graphs support users in deriving connections between different domains. These arguments and the requirements for bisociative knowledge discovery tools (from the previous section) lead us to the conclusion that a graph structure is the most promising approach for bisociative data exploration.

3.1 Tools for Data Exploration

The *Jigsaw* [12] system for investigative analysis across collections of text documents, the *Enronic* [18] tool for a graph based information exploration in emails (see Figures 2,3) and the *CET* [15,16] for efficient exploration and analysis of complex graph structures are some examples of exploration tools.

Since data exploration is an interdisciplinary topic at the intersection of knowledge discovery, visualization techniques and human-computer-interaction, we would like to structure the following overview into these subsections:

Data Set Analysis. By exploring data sets users want to locate anomalies, manage information and understand trends. It is very hard for users to deal with large, high-dimensional and heterogeneous data sets. Therefore, users need to be supported in their data analysis tasks by powerful data analysis methods [34]. Meanwhile, there exist a large amount of powerful algorithms for data analysis which deal with [14].

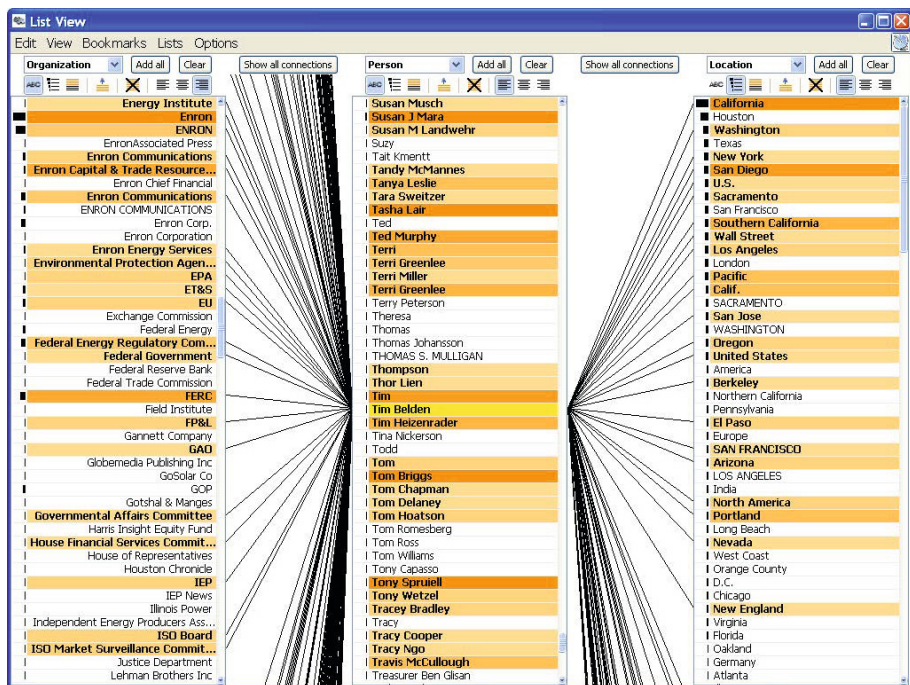


Fig. 2. List based visualization of an email dataset in the *Jigsaw* tool [12] showing connections of “Tim Belden”

Visualization Techniques Supporting Knowledge Discovery. Visualization, as a kind of external memory, supports users cognitive abilities [8] and enhances her or his memory while performing complex cognitive operations [28]. Tominski et al. [36] proposed an interactive graph visualization system *CGV*. This system is supposed to support users in visual exploration of linked information items. Ten years before Eick and Williams [11] already proposed a similar tool called *HierNet* for network-based visualization, which allows grouping, expanding and collapsing of information items. In *Jigsaw* [12] the authors present a system for investigative analysis across collections of text documents, exemplarily demonstrated on the ENRON data set. The same data set has been used in [18] to build up a graph based information exploration tool *Enronic*. While *ChainGraph* [23] is a further example for a graph-based exploration tool, designers of the *Tiara* tool [39] show that there are also other opportunities for supporting exploratory search than using graphs. Here, a visual summary based on different layers, which are organized in a coordinate system, is used to support exploratory search.

Human-Computer Interaction (HCI) for Explorative Tasks. Users interact with exploration tools in order to formulate problems and to solve them [32]. When considering HCI aspects in exploratory search, we need to take into account

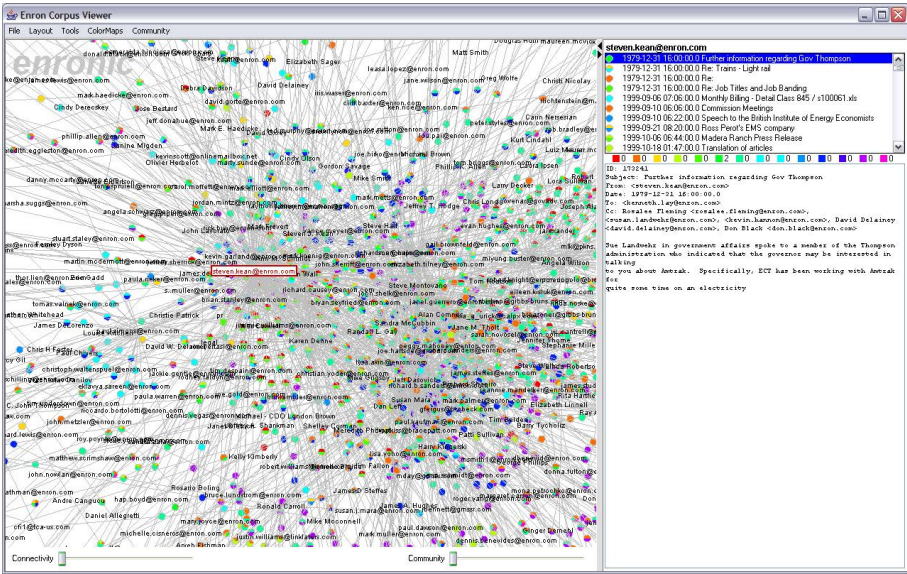


Fig. 3. Graph based visualization of an email dataset in the *Enronic* tool [18]

the fact that each new piece of information is providing the user with possible directions for further investigation, new insights and ideas [5].

4 Evaluation of Knowledge Discovery Tools

Usability evaluation is an integral part of user interface software development. With tool evaluation designers prove that their software fulfils its purpose and satisfies the needs of the target user, i.e. the software is “usable” [27]. With an evaluation it should be proven that, using exploration tools, users are able to make discoveries: effectively, efficiently and with positive attitude.

Our starting point are systems designed for exploration of large, heterogeneous and high-dimensional data sets. The research question that we target is how to evaluate such systems. The most important functionality of knowledge discovery tools is to support users in the creative discovery of new information and relations that were overlooked before in data sets. Thus, users of such tools usually have complex information needs.

Evaluation methods which can be used vary and consist of formal usability studies in the form of controlled experiments and longitudinal studies, benchmark evaluation of the underlying algorithms, informal usability testing and large-scale log-based usability testing [17]. There is also some research in the area of automatic evaluation of user interfaces [33]. Here the idea is using simulation to reflect the way a user is exploring a document collection. We consider an automatic approach, but it is not clear if this would work for biocreative exploration which requires creativity.

In the following, we discuss how to apply existing evaluation methods to exploration tools for biociative knowledge discovery. Evaluation of such complex systems is very challenging and requires collaboration with domain experts for creating scenarios and participation. Furthermore, complex information needs are usually vaguely defined and require much user time to be solved. In order to evaluate these tools more efficiently four components are essential: a standardized evaluation methodology, benchmark data sets, benchmark tasks and clearly defined evaluation measures [13].

4.1 Evaluation Challenges

Since knowledge discovery tools are complex systems [30], evaluation of them is very challenging. The first challenge is to create an appropriate scenario for evaluation. The tasks must be complex enough to represent a realistic situation. Such realistic exploratory tasks might require much time (sometimes weeks or even months) to be solved. Lab experiments are limited in time, therefore a “good balance” between time and the right level of complexity is crucial for lab user studies. Longitudinal studies, i.e. research studies that observe users’ usage of a system over long periods of time (e.g. months), overcome lab experiments drawbacks like strong time limitation and artificial environment. Researchers motivate the community to conduct long-term user studies because they can be well applied for studying the creative activities that users of information visualization systems engage in. [31]

Controlled lab studies and longitudinal studies require an involvement of target users. Unfortunately, knowledge discovery tools are often designed to be used by experts with domain-specific knowledge, e.g. molecular biologists, who are more difficult to find than participants without special skills or knowledge. Thus, the second challenge is recruiting the participants. This should be a group of people which represents the end users. It requires either collaboration with scientific institutions or some incentive (like money) to engage their participation [30]. In the study preparation step collaboration with domain experts is also needed to help the researchers in creation of appropriate scenarios.

4.2 Open Issues

By evaluating knowledge discovery tools we can either focus on the tool examination or carry out a comparative evaluation. Most researchers concentrate on evaluating their own tool to gain a deeper understanding of user interactions with it. However, the results do not provide such important information if or under what conditions their tool outperforms alternative tools for the same purpose. We found only one publication [20] that proposed an experimental design and a methodology for a comparative user study of complex systems.

To be able to compare and rank a tool among similar ones, benchmark data sets and tasks for user studies are essential [29]. Suppose we wanted to repeat the study conducted in [20] to compare our tool to theirs, we would need the document collection and the task solution used by the authors. However, this data is

not available to the public, so we cannot compare the results. A promising direction here is the *Visual Analytics Science and Technology* (VAST) contest¹ which offers data sets of different application domains with description and open-ended domain specific tasks. These tasks should be solved with the help of specific software within the contest. After the contest the solutions are made public, making the data available to evaluations. Thus, the data can be used for evaluations.

Additionally, clearly defined evaluation measures are also important in order to evaluate exploration tools more efficiently. These could be measures from different domains, e.g. information retrieval and human computer interaction, but new measures are still necessary in order to capture the amount of discoveries in document collections or how creative a solution is. The solution of a task itself can be very complex, so we need a way to account for answers which are only partially correct or complete.

The well established three usability aspects from HCI which are usually evaluated in user studies, are *effectiveness*, *efficiency* and *satisfaction* [1,17]. Each of these aspects can be expressed in various measures. In the context of discovery tools evaluation, one can express effectiveness in the amount of discovered information, efficiency in time to find new facts or in importance of the made discovery and satisfaction in the user's rating of the tool's comfort and acceptability [7]. All the three aspects should be ideally measured when evaluating a discovery user interface. Depending on the use case scenario, some of the criteria can be more important than others. If the exploration tool, for example, is primarily designed to support creative discovery of earlier unseen relations among data, the focus may lay more on user satisfaction and less on efficiency. It is not crucial to find the relations fast, but user's satisfaction by using the tool may directly influence his creativity, which is very important by bisociation discovery. A positive attitude helps the user to keep an open mind or play around with the information. User satisfaction, in general, is important, because if the user does not like the interface, one is not likely to use the tool any more again.

One can draw an analogy between user evaluation of exploration tools and automated benchmark evaluation of ranking algorithms in information retrieval. The latter requires a set of test queries, a document collection with labels according to relevancies (e.g. TREC) and a measure (e.g. Average Precision) [17], while discovery tools user evaluation requires a benchmark data set, a benchmark task with a standard solution and an evaluation measure.

4.3 Benchmark Evaluation for Discovery Tools

In the following we propose an evaluation method for discovery tools, consisting of two parts: The first part is a "small" controlled experiment with about 5–10 participants. The purpose of this is to collect qualitative data using user observations like audio/video recording and interviewing the participants afterwards. We actually do not need a special task to be solved by the participants. The assignment can be to discover new information using the software. From this

¹ <http://hcil.cs.umd.edu/localphp/hcil/vast11/>

study we collect data about learnability improvements and user satisfaction. We also get feedback about the users' favourite features and software drawbacks.

The second part is an online study, in which the software is provided to the participants as an online application. This makes it possible to overcome the time limitation found in lab experiments. The participants can access the tool from their own working environment and spend as much time as they like with the tool, even working discontinuously. After that they can use an online questionnaire to provide the task solution and usability feedback. Participants are motivated to solve an interesting task using the tool. We assume that the VAST benchmark data with an investigative task (from IEEE VAST 2006 Contest) can be used as a benchmark data set and a benchmark task. The tool interactions of each participant are logged on the server side. Each participant can spend arbitrarily much time to solve the quest.

We can analyze the log files to get the time spent by participants, to get the solution and interaction patterns. The outcome of the study also contains the number of participants who succeeded in solving the task in comparison to all participants who tried. Each participant is motivated to answer an online questionnaire to provide the task solution and usability feedback. It is beneficial to get the user feedback during the study as it may forget some important issues due to the extended duration. This can be done in the form of a diary. The purpose of the second part is to collect quantitative data.

The described method is only the first step in the creation of a good methodology. It still has several limitations. The first problem is to get an appropriate number of participants. It is not easy to stimulate the participation even with money and if it would work the study becomes cost consuming. One possible solution lies in automatic evaluation (see, e.g., [4]). We could simulate exploration process on different levels and for diverse tasks. However it is not clear how to model a *creative* exploration process, which is important in the case of tasks like creative information discovery, e.g. of relations that were overlooked before in data sets. We also do not have a clear understanding how to judge the success of the search given a complex information need. Thus, the question about evaluation measures remains.

5 Conclusion and Future Work

In this chapter we tried to elaborate on a definition for the term *data exploration*, which—although often used—has no formal description. Our findings, however, can only serve as a starting point for more thorough research on contexts and tools which are recognized as *exploratory* and therefore should be covered by a formal definition of the term. A selection of tools for graph and data exploration has been provided in Section 3. We propose a simple classification, however creating a taxonomy on these tools would on the one hand require a more specific context, on the other hand we would need a definitive formal specification of what exactly an exploratory tool would be. Finally, we have discussed the problem of evaluating the performance and usability of exploratory tools and

identified two main issues: First, evaluation scenarios for those tools are much more complex, resulting in longer sessions and more effort. Second, there is no benchmark against which a tool can be tested, i.e. each evaluation needs to come up with a reference scenario of itself.

With the VAST data set we proposed a starting point for finding such benchmark, which still needs to be specified and tested. However, several aspects are yet unclear. This applies to evaluation methodology, in particular the possibility to evaluate the discovery tools automatically, and evaluation measures. We would like to motivate the community and make the researchers pay attention to the fact that evaluation of knowledge discovery tools should be carried out using a standardized evaluation methodology in combination with benchmark data sets, tasks and measures. Only then discovery tools designers can evaluate their tools more efficiently.

Acknowledgement. The work presented here was supported by the European Commission under the 7th Framework Programme FP7-ICT-2007-C FET-Open, contract no. BISON-211898 and the German Ministry of Education and Science (BMBF) within the ViERforES II project, contract no. 01IM10002B.

Open Access. This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. ISO 9241-11: Ergonomic requirements for office work with visual display terminals (VDTs). Part 11 - guidelines for specifying and measuring usability. Geneva: International Standards Organisation. Also available from the British Standards Institute, London (1998)
2. Collins English Dictionary - Complete and Unabridged. HarperCollins Publishers (2003)
3. The American Heritage Dictionary of the English Language. Houghton Mifflin (2009)
4. Azzopardi, L., Järvelin, K., Kamps, J., Smucker, M.: Proc. of SIGIR 2010 Workshop on the Simulation of Interaction: Automated Evaluation of Interactive IR (SimInt 2010). ACM Press (2010)
5. Bates, M.J.: The design of browsing and berrypicking techniques for the online search interface. *Online Review* 13(5), 407–424 (1989)
6. Berthold, M.R. (ed.): *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250. Springer, Heidelberg (2012)
7. Bevan, N.: Measuring usability as quality of use. *Software Quality Journal* 4(2), 115–130 (1995)
8. Card, S., Mackinlay, J., Shneiderman, B.: *Using visualization to think*. Readings in Information Visualization (1999)
9. Dengel, A., Agne, S., Klein, B., Ebert, A., Deller, M.: Human-centered interaction with documents. In: *Proceedings of the 1st ACM International Workshop on Human-centered Multimedia, HCM 2006*, pp. 35–44. ACM, New York (2006)

10. Dubitzky, W., Kötter, T., Berthold, M.R.: Towards Creative Information Exploration Based on Koestler's Concept of Bisociation. In: Bisociation, Part I. Springer (2011)
11. Eick, S.G., Wills, G.J.: Navigating large networks with hierarchies. *Readings in Information Visualization Using Vision to Think*, 207–214 (1999)
12. Görg, C., Stasko, J.: Jigsaw: investigative analysis on text document collections through visualization. In: *DESI II: Second International Workshop on Supporting Search and Sensemaking for Electronically Stored Information in Discovery Proceedings*. University College London, UK (2008)
13. Gossen, T., Haun, S., Nuernberger, A.: How to Evaluate Exploratory User Interfaces? In: *Proceedings of the SIGIR 2011 Workshop on "Entertain Me": Supporting Complex Search Tasks*, pp. 23–24. ACM Press (2011)
14. Han, J., Kamber, M.: *Data mining: concepts and techniques*. Morgan Kaufmann (2006)
15. Haun, S., Gossen, T., Nürnberger, A., Kötter, T., Thiel, K., Berthold, M.R.: On the Integration of Graph Exploration and Data Analysis: The Creative Exploration Toolkit. In: Berthold, M.R. (ed.) *Bisociative Knowledge Discovery*. LNCS (LNAI), vol. 7250, pp. 301–312. Springer, Heidelberg (2012)
16. Haun, S., Nürnberger, A., Kötter, T., Thiel, K., Berthold, M.R.: CET: A Tool for Creative Exploration of Graphs. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) *ECML PKDD 2010*. LNCS, vol. 6323, pp. 587–590. Springer, Heidelberg (2010)
17. Hearst, M.: *Search user interfaces*. Cambridge University Press (2009)
18. Heer, J.: Exploring Enron: Visualizing ANLP results (2004), <http://hci.stanford.edu/jheer/projects/enron/v1/>
19. Herczeg, M.: *Interaktionsdesign*. Oldenbourg Wissenschaftsverlag (2006)
20. Kang, Y., Goerg, C., Stasko, J.: How can visual analytics assist investigative analysis? design implications from an evaluation. *IEEE Transactions on Visualization and Computer Graphics* (2010)
21. Kreuzeler, M., Nocke, T., Schumann, H.: A history mechanism for visual data mining. In: *Proceedings of the IEEE Symposium on Information Visualization (Infovis 2004)*. IEEE Computer Society (2004)
22. Leinhardt, G., Leinhardt, S.: Exploratory data analysis: New tools for the analysis of empirical data. *Review of Research in Education* 8, 85–157 (1980)
23. Lohmann, S., Heim, P., Tetzlaff, L., Ertl, T., Ziegler, J.: Exploring Relationships between Annotated Images with the ChainGraph Visualization. In: Chua, T.-S., Kompatsiaris, Y., Mérialdo, B., Haas, W., Thallinger, G., Bailer, W. (eds.) *SAMT 2009*. LNCS, vol. 5887, pp. 16–27. Springer, Heidelberg (2009)
24. Marchionini, G.: Exploratory search: From finding to understanding. *Communications of the ACM - Supporting Exploratory Search* 49(4) (April 2006)
25. Miller, G.A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Science* (63), 81–97 (1956)
26. Naisbitt, J.: *Megatrends. Ten New Directions Transforming Our Lives*. Warner Books (1982)
27. Nielsen, J.: *Usability engineering*. Morgan Kaufmann (1993)
28. Norman, D.: *Visual representations. Things that Make us Smart: Defending Human Attributes in the Age of the Machine* (1994)
29. Plaisant, C.: The challenge of information visualization evaluation. In: *Proceedings of the Working Conference on Advanced Visual Interfaces*, pp. 109–116. ACM (2004)

30. Redish, J.: Expanding usability testing to evaluate complex systems. *Journal of Usability Studies* 2(3), 102–111 (2007)
31. Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In: *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*, pp. 1–7. ACM (2006)
32. Spence, R.: *Information Visualization*. Addison Wesley/ACM Press, Harlow (2001)
33. Stober, S., Nürnberg, A.: Automatic evaluation of user adaptive interfaces for information organization and exploration. In: *SIGIR Works. on SimInt 2010*, pp. 33–34 (July 2010)
34. Thomas, J.J., Cook, K.A.: *Illuminating the path –The research and development agenda for Visual Analytics*. IEEE Computer Society Press (2005)
35. Tominski, C.: *Event-Based Visualization for User-Centered Visual Analysis*. PhD thesis, Institute for Computer Science, Department of Computer Science and Electrical Engineering, University of Rostock (2006)
36. Tominski, C., Abello, J., Schumann, H.: Cgv—an interactive graph visualization system. *Computers & Graphics* 33(6), 660–678 (2009)
37. Tukey, J.: *Exploratory data analysis*. Addison-Wesley, Massachusetts (1977)
38. Tukey, J.W.: *Exploratory Data Analysis*. Addison-Wesley (1977)
39. Wei, F., Liu, S., Song, Y., Pan, S., Zhou, M.X., Qian, W., Shi, L., Tan, L., Zhang, Q.: Tiara: A visual exploratory text analytic system. In: *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2010*, pp. 153–162. ACM, New York (2010)