Abstract

Several information visualization techniques have been developed in the last years due to the need of representing and analyzing the huge amount of data generated by several applications or made available through the World Wide Web. These techniques are usually interactive and provided as part of a graphical user interface. Information visualization techniques are usually reported showing their use in experimental situations, employing some kind of analysis. Recently, studies have specifically addressed the evaluation of such techniques. This paper reports results from the definition and application of specific criteria for the evaluation of information visualization techniques. The criteria encompass the evaluation of both visual representations and interaction mechanisms, within the framework of usability testing methods.

CR Categories: I.3.6 [Methodology and Techniques]: Interaction Techniques; Ergonomic; H.5.2 [User Interfaces]: Evaluation/Methodology

Keywords: evaluation, ergonomic criteria, usability testing

1 Introduction and motivation

In the last few years the increasing volume of information provided by several applications, different instruments and mainly the Web has lead to the development of techniques for selecting among a bulk of data the subset of information that is relevant for a particular goal or need. Research on visual query systems, data mining and interactive visualization techniques has resulted in a wide variety of visual presentation and interaction techniques that can be applied in different situations.

However, although there is a great variety of models and techniques for information visualization [Card et al. 1999] [Spence 2001], each application requires a particular study in order to determine if the selected technique is useful and usable. These studies are usually guided by the type of data that should be represented and the user tasks or analysis process that the visualization should help or support. It is clear that we cannot separate the visual aspects of both data representation and graphical interface from the interaction mechanisms that help a user to browse and query the data set through its visual representation. Moreover, evaluating these two aspects is an important issue that must be addressed with different approaches of usability evaluation including, for sure, experimental studies with users.

Usability is a term employed to describe the quality of use of applications by end-users [Bevan 1995]. In the context of interfaces for information visualization users not only interact with widgets on the interface but also with data supporting decision-making, which could be affected by the way information is presented. In general, information must be cut and summarized to be useful for supporting decision-making, even though the kind of information processing could alter the quality of original data set. These problems are not related to the interaction mechanisms provided by the interface but to the data processing itself. Although stability in data visualization, i.e., minimal change from the result of an action to the next one, would help users in their analysis process, the actual visual representation of information plays an important role in the usability of data. For example, when presenting n-dimensional data in a 2D or 3D visualization, techniques must adapt information coding to visualization constraints such as the reduction of dimension.

Since information visualization is intended to provide insight from data, it becomes clear that both visual representation and interaction techniques must not affect the ways the user needs to use the data in a variety of analysis procedures. We separate usability issues in three main categories: i) visual representation usability, referring to the expressiveness and quality of the resulting image; ii) interface usability, related to the set of interaction mechanisms provided to users so they can interact with data through the visual representation; and iii) data usability, devoted mainly to the quality of data for supporting users’ tasks. In this paper we will concentrate on visual representation and interface usability issues.

Our approach is to link interface usability knowledge, concepts and methods (typical from HCI – human-computer interaction domain) with evaluation of the expressiveness, semantic content, and interaction facilities of visualization techniques. Based on the classical ergonomic criteria from Bastien and Scapin [1993], we have devised specific criteria for the evaluation of visual representations and interaction mechanisms provided by information visualization techniques. At this stage of the work, we concentrate on usability evaluation of hierarchical information visualization techniques.

The paper is organized as follows. Section 2 presents a further discussion on usability methods and related work on the evaluation of information visualization techniques. Section 3 describes our set of criteria for evaluating hierarchical information visualization techniques, whereas section 4 reports a case study on the evaluation of a specific visualization technique. Finally, section 5 discusses our contribution and concludes with comments on how to advance towards the adaptation of HCI usability methods to information visualization techniques.
Usability evaluation methods have been developed to test and/or measure the efficiency, interaction flexibility, interaction robustness, and quality of use of user interfaces. Most methods are based on experiments with users [Rubin 1994] where usability is measured by observation of users interacting with the interface as well as data about task completion time and errors during execution of tasks. Other usability evaluation methods are based on the inspection of interfaces by an expert, which is able to recognize usability problems [Nielsen and Mack 1994]. This kind of inspection can be based on heuristics like Nielsen’s [Nielsen 1993] or the ergonomic criteria from Bastien and Scapin [1993].

The main aim of usability evaluation is to identify problems that interfere on users’ tasks, causing stress or reducing user performance. These techniques are quite efficient for evaluating usability in interfaces when concrete tasks are considered. For example, in a heuristic evaluation following either Nielsen’s heuristics or Bastien and Scapin’s criteria, an HCI expert experiments with the system reproducing its intended use; when a usability problem is found, it is rated according to severity levels and classified under one or more criteria. At the end of the inspection, system developers may use this list to guide improvements in the interface. Equally important results can be obtained from experiments with users, where tasks execution might reveal the impact of usability problems in user performance.

However, it is much harder to evaluate usability when abstract tasks such as “understand data” or “make decision based on information” are considered. In addition, interfaces for information visualization include a set of 2D and 3D structures (such as 3D objects, polygons, scenarios and virtual worlds) that are unusual in most WIMP interfaces. As a consequence, it is much more difficult to describe and to identify usability problems on this kind of interface than in WIMP ones. The absence of specific criteria for evaluating information visualization interfaces is a great barrier since most metrics used to evaluate usability such as accomplishment of tasks and user performance might be less important for such interfaces if we consider the effective usage of information as the most important goal.

Despite the importance of such data usability measure, information visualization techniques were introduced in the past without reporting evaluation results (see, for example, Cone Trees, Treemaps and Hyperbolic Browser first descriptions [Robertson et al. 1991][Lamping et al. 1991] [Johnson and Shneiderman 1991]). Further works have addressed the comparison of techniques showing their use in experimental situations, usually with (few) users testing [Carriere and Kazman 1995][Wiss et al. 1998][Wiss and Carr 1999][Pirolli et al. 2000].

Recent reports have shown the growing importance given to the evaluation of visualization techniques [Chen and Czerwinski 2000][Barlow and Neville 2001][Kobza 2001]. The observation of the reported evaluations reveals that the usual way to evaluate information visualization techniques still is experiments with users measuring task completion time and answer correctness. Evaluations explicitly involving cognitive aspects or semi-formal HCI methods have been addressed by few authors. For example, the naming time method was used to evaluate the effect of the reduction of quality of 3D images and the quality of information provided, demonstrating the use of some cognitive aspects of visual representation quality and user performance related to the time spent for identifying objects and understanding information [Watson et al. 2000]. Pirolli and Rao [1996] used GOMS – Goals, Operators, Methods, and Selection Rules to evaluate Table Lens for exploratory data analysis tasks.

The convergence of information visualization and HCI evaluation methods has built the adequate scenario to discuss specific criteria for the evaluation of such techniques. Specific criteria were addressed in few works along these years. Brath [1997] proposed specific measures for static visualizations and Yang-Pelaez and Flowers [2000] discussed the information content in visual representations. Morse et al. [2000] centered the evaluation of visualizations provided by information retrieval systems in a subset of the tasks defined by Zhou and Feiner [1998], which in turn have their origin in the work of Wehrend and Lewis [1990]. Our approach contributes to this scenario by providing the insertion of specific criteria for the evaluation of usability of information visualization techniques in the context of both the ergonomic criteria from Bastien and Scapin and Nielsen’s heuristics.

### 3 Criteria for evaluating information visualization usability

The evaluation of information visualization techniques should be based in testing both visual representation and interaction mechanisms. For example, usual and critical aspects of visual representations are object occlusion and visual disorder, while visual disorientation is caused by changes in the visual representations due to some user action. Thus, there are (many) situations when one aspect (interaction) affects the other (visual representation). All such characteristics should be verified in order to evaluate a specific visualization technique.

Our approach is based on two sets of criteria: the first being for usability testing of visual representations and the second one, for evaluating interaction mechanisms. Two initial sets were defined in a previous work [Freitas et al. 2002] based on case studies with different visualization techniques. Then, those two sets were refined taking into account characteristics found mainly in information visualization techniques, thus excluding many scientific data visualization techniques. As a last refinement, these criteria were tailored for hierarchical information visualization to be validated with well-known techniques. The resulting two sets are examined in the following sections.

#### 3.1 Visual representation criteria

Figure 1 presents a diagram depicting the criteria for evaluating visual representations. They are commented in the following paragraphs along with examples of metrics focusing hierarchical information, when appropriate.

The first (and main) aspect to be evaluated in a visualization technique is the codification of information. Perception of information is directly dependent on the mapping of data elements to visual objects. This can be enhanced by the use of additional symbols or realistic characteristics. Another important aspect is the use of alternative visual attributes or objects to represent information derived from the data like groups of elements in clustered representations.

Although information coding plays an important role for the cognitive complexity of an image, other aspects can be used to evaluate this characteristic: data density, data dimension and the
relevance of the displayed information. Within hierarchical information visualization, data density is related to the number of data elements presented to the user, while data dimension corresponds to the number of hierarchical levels in the visualization. Regarding the relevance of displayed information, this criterion allows verifying if the technique provides means to distinguish relevant information from the subset of data represented in the display. An example would be having a special representation or visual characteristic for the root node of the hierarchy.

Spatial organization is related to the overall layout of a visual representation, which comprises analyzing how easy it is to locate an information element in the display and to be aware of the overall distribution of information elements in the representation. Locating an information element can be hard if some objects are occluded by others, and if the layout does not follow a “logical” organization depending on some characteristics of the data elements. So, degree of object occlusion and logical order are characteristics to be measured in the visual representation, concerning to the location of objects. The spatial orientation, which contributes for the user being aware of the distribution of information elements, is dependent on the display of the reference context while showing details of one or more specific elements.

The semantic contents of data to be displayed may also be affected by limitations in the visual representation, which are geometric or visual constraints like maximum number of different geometric objects or symbols, and maximum number of data elements or levels in a hierarchy, which may be imposed by the technique or the implementation. The evaluation of a technique under this criterion has the goal of identifying both the origin of the limitation (technique or implementation) and its effect.

Finally, an important aspect of information visualization techniques is the result of rebuilding part or the entire visual representation after a user action. The transition between states may cause different usability problems due to the time the technique takes to re-display information as well as the spatial organization of the resulting image which may reflect on the sense of orientation.

### 3.2 Interaction mechanisms criteria

Information visualization techniques should support different user’s task depending on the application domain, and different, but overlapping sets of such tasks are described in the literature [Wherend and Lewis 1990][Shneiderman 1996][Zhou and Feiner 1998].

Taking into account these tasks, we can identify interaction mechanisms that are may be provided by visualization techniques depending on their goal. To support operations on a hierarchy, techniques may provide high-level mechanisms such as searching, filtering, pruning, clustering and expansion. We can say that these mechanisms provide means of interacting directly with the data set. On the other hand, there are some low-level mechanisms that are based on interaction with the visual representation: selection, viewing transformation, geometrical manipulation, zooming and re-display of previous representations. Together, both classes of interaction mechanisms provide ways to interact with the information space.

The usability of an interactive tool is commonly evaluated using a task model and Nielsen’s usability factors [Nielsen 1993] regarding learning and memorization, efficiency of use (task completion time), user’s subjective satisfaction and user’s errors during task completion. Within information visualization techniques we are more concerned with data usage effectiveness; thus, aspects like learning and memorization of system’s functions are less important. Our proposal is to use only four criteria for the evaluation of interaction mechanisms: easy of use, efficiency (task completion time), user’s subjective satisfaction and user’s errors.

![Figure 1: Criteria for visual representation evaluation.](image)
3.3 The Criteria and Usability Testing Methods

Although we provided a set of criteria, its use in a practical evaluation should be performed under the framework of a usability testing method.

Depending on the purpose of the evaluation, different methods may be used with these criteria. Heuristic evaluations of a prototype by a group of experts would collect many usability problems. In this situation, the criteria would be translated to heuristics, i.e., to characteristics that should be verified by the expert. On the other hand, conceiving tasks for experiments with users would provide more controlled ways to verify the effect of the visual representation and the interaction mechanisms in user performance (time and errors).

We have investigated the benefits of our criteria by means of a case study where their use in a heuristic evaluation of an information visualization technique was compared to the use of the traditional Nielsen’s and Bastien and Scapin’s sets. The following section describe this case study.

4. Case study: Evaluation of the Bifocal Browser

In the process of refining the criteria described in this paper, four visualization techniques were evaluated as preliminary case studies: the Magnifind\(^1\) implementation of the hyperbolic browser, CHEOPS [Beadouin et al. 1996], Treemaps [Johnson and Shneiderman 1991] and an in-house tool, the Bifocal Browser [Cava and Freitas 2001]. For the experiment reported herein, however, we used only the Bifocal Browser. In the following subsections, we shortly describe the Bifocal Browser, and then we show the results of its evaluation using the three sets of heuristics.

4.1 Bifocal Browser: a short description

The Bifocal Browser [Cava and Freitas 2001] is an alternative way to explore large hierarchies in a node-edge diagram, incorporating some features borrowed from space-filling approaches [Johnson and Shneiderman 1991; Andrews and Heidegger 1998] and the hyperbolic browser [Lamping et al. 1991]. The technique is based on the focus + context approach, and uses two foci instead of one. In this technique, the hierarchy is represented as a node-edge diagram separated in two connected sub-diagrams, defining separate areas in the window: a detail area, which shows the node of interest and its subtree, and a context area, that displays its parent and siblings subtrees. Although it is not based on the hyperbolic geometry, the technique has a radial layout. In Figure 2, the hierarchy is anchored on two main nodes called the context and detail focus, displayed at the center of the context and detail areas, respectively. The central rectangle in the right (Figure 2b) represents the detail focus at a certain moment, i.e., a node of interest, while the left one (Figure 2a) is the context focus. Thus, at the same time that it provides information about hierarchical relationships, the technique shows a detailed view of the subtree containing the node of interest.

Both context focus and detail focus are located side by side separated by an arbitrary (and parameterized) distance, defining separate circular areas in the window. Once the user indicates a node as the point of interest, the whole subtree of this node is shown in the right area of the window, while its parent node and all other subtrees are displayed in the left half of the window. Each subtree is displayed in a radial layout, with the selected node at the center of a circle, and its descendants distributed in concentric circles depending on their level in the structure. In order to avoid occlusion among objects, each subtree is actually displayed in a circle sector.

Nodes are displayed as rectangles with the size depending on their location in relation to the focus point. A node that is distant from the focus is shown with less detail than nodes nearer the focus, while nodes beyond a certain distance are not shown. This strategy was adopted to avoid displaying and manipulating elements that are far from the point of interest, based on the idea that a user browses a structure until reaching a specific node. Moreover, due to the reduced size these nodes would have if displayed near the border of the circle, the user probably would not point at them.

A different color is used to display the subtree that has recently occupied the detail area. Also, the root node is always displayed in red, in order to keep the user aware of what level in the hierarchy is currently in the context focus. This intends to minimize a possible disorientation that might happen due to rotations and translations applied to the subtrees when the user selects another node.

Distribution of nodes around the root node takes into account the number of leaf nodes in each subtree. Therefore, the subtree with more leaf nodes occupies the largest sector in the circle. This rule is applied recursively to each subtree in the structure. On the other hand, in the detail area, the goal is to provide the representation of the hierarchical relationships by means of a tree with the interest node as root. To achieve this goal, nodes at the first level are uniformly distributed around the inner circle. Their subtrees, however, are displayed in sectors with size proportional to the number of leaf nodes, following the same rule applied in the context area. This difference in the layout is more evident in unbalanced trees.

The selection of a node is the main interaction mechanism provided to the user; it can be applied to any node in order to bring to the detail focus the node along its subtree. This operation imposes a translation of the subtree to the detail area, as well as a rotation in the structure displayed in the context area.

---

\(^1\) Now available as the Star Tree (http://www.inxight.com)
4.2. Evaluating the Bifocal Browser

The Bifocal Browser was evaluated by a group of 14 computer science students at the end of a course on usability inspection methods and three experts. Before the actual evaluation, all the students and the experts had a training session, during which they played freely with the browser without interference.

The students were divided into three groups, one for each set of heuristics. Thus, four students evaluated the Bifocal Browser employing a heuristic evaluation with Nielsen’s criteria, five with Bastien and Scapin’s ergonomic criteria, and five with our set. A different set of criteria was assigned for each expert, to serve as a control in relation to the students.

For the evaluation, they received specific instruments with the description of each heuristic and a form-like document to record the usability problems under one or more heuristics, assigning a severity level (from 0 – no important to 4 – catastrophic) to each one. They performed the evaluation using two hierarchies, with 300 and 1000 nodes respectively, during one hour and a half. These hierarchies represent fake road names in order to guarantee that no previous knowledge of structure exists.

Results from the evaluation performed by students show the detection of 39 problems with Nielsen’s criteria (Figure 3), 46 problems with Bastien and Scapin’s ergonomic criteria (Figure 4) and 52 problems with our set (Figure 5).

![Figure 3: Results from the evaluation using Nielsen’s heuristics.](image)

![Figure 4: Results from evaluation using Bastien and Scapin’s ergonomic criteria.](image)

Students using Nielsen’s heuristics as well as Bastien and Scapin’s ergonomic criteria found much more problems in the GUI (graphical user interface) than in the visualization technique (layout and interaction). Similar evidences are found observing the results from the evaluation by the experts. The same was not true for the students and expert using our set of criteria: they found much more problems in the technique than with the GUI (Figure 6).

![Figure 5: Results from evaluation using our criteria.](image)

![Figure 6: Comparison between finding of GUI-related problems and visualization technique usability problems (in %).](image)

Many of the detected problems can be explained by the fact that the Bifocal Browser was implemented only to test the Bifocal Tree representation [Cava et al. 2002], without any concern regarding GUI design like buttons, error messages or help. However, it should be noticed that even though our criteria were not defined to detect traditional GUI problems, a considerable number of them was found during the evaluation. Another important result was that evaluators who employed our set of criteria detected 31 problems with severity levels 3 and 4, while those using Nielsen’s and Bastien and Scapin’s sets found 14 and 22, respectively.

5. Discussion and Final Comments

Evaluating user interfaces is usually accomplished to detect design problems in the layout as well as during interaction. In information visualization techniques, interface usability issues are intertwined with the expressiveness of the visual representation. These two aspects are equally important and have to be evaluated in order to verify how well a visualization technique supports users’ tasks.
As can be seen from the literature, only recently usability inspection methods, mainly experiments with users, have been undertaken to effectively evaluate information visualization techniques [Chen and Yu 2000]. Usability experiments have been carried out to verify completion of tasks and difficulties in interaction using selected visual representations or selected tools. Their goal, however, was not to set a framework for evaluating visualization techniques but to specifically test design features adopted in different visual representations or different systems functions.

Our approach for evaluating information visualization techniques considering a specific set of criteria, categorized by visual representation characteristics and usability factors in interactive tools, addresses different aspects than those reported in recently published literature. Besides facilitating a deeper evaluation of an information visualization technique regarding its visual representation, the close relationship among our criteria and the classical ones from usability inspection methods allows us to identify also GUI-related problems.

The case study, although brief, demonstrates the benefits of our criteria as an important aid to the task of evaluating information visualization techniques. The main problems detected in the visual representation of the Bifocal Browser were the occlusion of nodes in the context area in large hierarchies and the disorientation produced by the change in the overall layout, when one selects a different node as the interest node. Although the first one is a well-known problem in node-edge diagrams, pruning and clustering minimizes it, and are provided by the browser. The disorientation associated to state transition due to lack of animation in the current implementation was also evident at the first interaction, and was reported by all evaluators. Moreover, some problems known only by the developer were also detected by a student using our criteria. Regarding interaction mechanisms, some features like filtering are missing, and the nodes are only distinguished by their names. An additional attribute could for sure be associated with color (for example, file extension, if the hierarchy is a file tree) but this is not available in the current implementation.

The comparison with evaluations employing different sets of heuristics allowed us to verify the expressiveness (or completeness) of the proposed set, concerning to usability of the visual representation and related interaction mechanisms. However, it should be noticed that this set has been tailored for the evaluation of hierarchical information visualization. The results showing the report of GUI-related problems allow us to affirm that our set has much coverage than we intended to, and further evaluations must be made to test this evidence.

Finally, another important result for the integration of usability testing methods and evaluation of information visualization techniques is the flexibility of our set. We have employed our criteria with other testing methods, namely with conformity inspection and user’s experiments. For the conformity inspection we need only to adapt the description of criteria to an instrument containing a checklist where each criterion was mapped to one or more conformance questions. For experiments with users, tasks were designed for each interaction mechanism in order to collect data about task completion time, users’ errors and satisfaction. Results from the evaluation of the Bifocal Browser by users were reported elsewhere [Cava et al. 2002].

Next step already being planned is to perform case studies with information visualization techniques that follow different paradigms like the space-filling approach, as well as with non-hierarchical information visualization techniques. An interesting issue that might also be addressed is to develop metrics for each criterion in order to include some of them in monitoring tools.

References


Acknowledgments

This work has been sponsored by CNPq, FAPERGS and CAPES. We gratefully acknowledge discussions with M. Winckler, M. Pimenta, C. Farenc and E. Diniz. We also are grateful to the students and experts who participated in the evaluations.