# THE USE OF GUIDELINES TO OBTAIN USABILITY FOR GEOGRAPHIC INFORMATION INTERFACES


Benedikt Heil
WSL Institute for Snow and
Avalanche Research SLF
Flüelastrasse 11
7260 Davos Dorf
Switzerland
heil@slf.ch

Tumasch Reichenbacher
Dept. Of Geography
University of Zürich
Winterthurerstrasse 190
8057 Zurich
Switzerland
tumasch.reichenbacher@geo.uzh.ch



## Abstract

This paper addresses the applicability of classical Human Computer Interaction (HCI) methods to design usable interfaces to geographic information and the question for a need of specific methods taking into account the characteristics of spatial data. Based on design guidelines three prototypes of a user interface to an information platform for natural hazards were designed and evaluated in an experiment using standard usability evaluation methods and eye movement recordings. The results show that the design solution for an interface with a balance of thematic navigation inside a tree and map-based, spatial navigation scored best. It is therefore suggested to harness a combination of HCI methods and an independent design theory for spatial applications to develop usable geographic information interfaces.


## Introduction

In the last decade digital screen-based mapping tools have become common for visualizing spatial data in professional and public settings. The major advantages of digital maps over paper maps are their flexibility stemming from the separation of information storage and display, the possibility of adding interaction, the possibility of depicting actual data, and the ease of disseminating geographic information to a large audience. Today, many sophisticated and highly interactive geographical information systems are accessible through the Internet. The usability of their graphical user interfaces (GUI) varies greatly. The major usability problems seem to have their seeds in the limited screen size, the lack of spatial overview, i.e. context, the problems associated with the fact that a lot of data can be depicted at the same time, the large choice of functions offered, and the heterogeneous user groups in terms of experience with computer use and training with the system itself. Geovisualization scientists and cartographers have therefore become interested in borrowing the insights of Human Computer Interaction (HCI) and usability approaches to make sure that their systems are usable before their release (Slocum et al., 2001). However, many researchers questioned if it is adequate to employ HCI theory and methods to solve the usability issues in geographic information technology and systems, claiming special and unique characteristics of spatial data and systems that require special methods. Hence, the need for an independent theory to support methods for usable geovisualization has been identified as an area requiring more detailed research (see for instance MacEachren and Kraak, 2001, Fuhrmann et al., 2005). Similarly, a shift in the design process for geovisualizations towards a user-centred design has been postulated (e.g. Nivala et al., 2008). Additionally, Fabrikant (2005) states that there is still relatively little known about the effectiveness of interactive graphical representations and geovisualization tools.

Besides classical methods to determine the effectiveness of GUIs like measuring the time or number of mouse clicks needed to solve a certain task, an extended way is observing the user's eye movements during the usage of a system (Çöltekin et al., 2009). The overall assumption in regarding eye movement data is that more fixations of the user's gaze on the screen indicate a higher amount of visual processing that is needed to understand the content of the interface (Duchowski, 2007). It can therefore be presumed that fewer fixations indicate a more comprehensible and effective interface.

In a case study accompanying the development of a user interface for a spatial information system, a formative evaluation was conducted involving experiments with potential future users. This system developed by the WSL Institute for Snow and Avalanche Research SLF in collaboration with the Federal Office of Meteorology and Climatology (MeteoSwiss) and the Federal Office for the Environment (FOEN) is a joint information platform for natural hazards (Gemeinsame Plattform Naturgefahren, in short GIN) for Switzerland. It will integrate current and recent meteorological, hydrological and snow data, as well as up-to-date observations and forecasts. The objective of the platform is to offer experts, government executives in charge and action forces (e.g. army, civil protection, fire brigades etc.) an effective and efficient tool to retrieve all information relevant to natural hazards during an event and get an overview of the situation in a coherent and unified way. The database accessed by GIN holds measurements from gauging stations (e.g. water level), observations (e.g. avalanche releases), and forecasts (e.g. runoff prediction). This data is presented as maps, tables, diagrams, and text. Temporal analysis of the past development of a parameter is also foreseen. At the time of writing the provisional GUI is being adjusted according to the results of this study and the feedback of the test users.

For the reminder of this paper we sketch the goals of the study and describe the methods applied including a description of the user experiment. Further, we present the results of that experiment, discuss them, and draw conclusions.

## Objectives

The overall objective of our research was to provide better insight into the design of usable interfaces for geographic information platforms, addressed in two sub goals.

The first goal was to ascertain if the body of existing visualization guidelines and the theories of human computer interaction are directly applicable to the kind of geographic information platforms described above. The intention was to review design guidelines from HCI research and establish design alternatives based on them for the interface of the GIN platform. A second goal was to find the most appropriate, i.e. effective and usable design solution among these alternatives that can serve as a reference for implementation.

## Methodology

In order to address the above mentioned objectives and to assess the usability of geographic information platforms in general and the GIN platform in specific based on a review of existing design guidelines and best practice as well as a study of existing web mapping and web Geographic Information Systems (GIS), three possible GUI drafts for GIN were developed. These were evaluated regarding their effectiveness in an early development phase. They were realised as mock-ups with limited functionality and implemented in Adobe Flash. The design alternatives differ in their employment of a hierarchical navigation bar and their degree of using spatial navigation strategies, i.e. a map-based navigation. The following paragraphs and Figure 1-3 give an overview on the specific functional elements of the drafts and emphasize the differences between them.

### *The E-Draft*

The E-Draft (E stands for Explorer) was designed following the well known structure of the Microsoft Windows Explorer and is therefore not based on a specific design guideline. Data was organized in a navigation tree along the left border of the interface (see Figure 1). The whole thematic navigation only takes place inside the tree. Since all available information could be found in the navigation tree, its content needed to be extremely detailed. Navigation to a gauging station at a specific location had therefore to be accomplished completely inside the tree, resulting in a very long tree while navigating. Due to the fact that detailed thematic navigation was performed in the tree, the maps were static and did not provide any functionality. This draft was meant to represent a non-spatial navigation and interaction strategy for the interface. However, a difference to the Windows Explorer was the grouping of data in tabs to enable a sorting of the depiction style (e.g. maps, diagrams and tables).

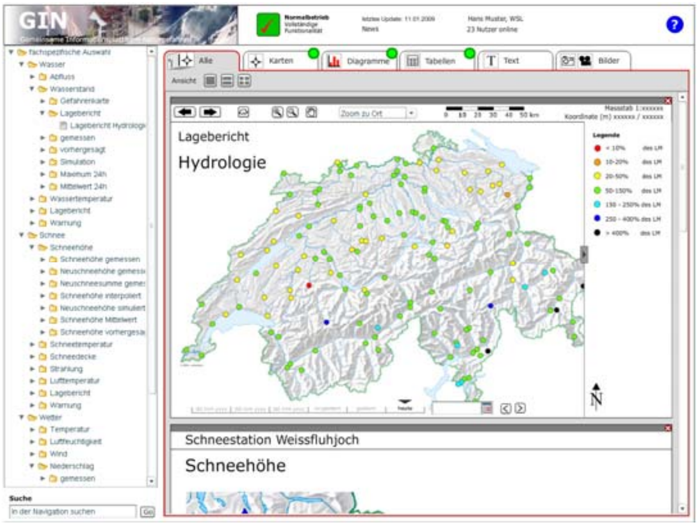


Figure 1. The E-Draft based on the MS Windows Explorer navigation bar

*The G-Draft*

The G-Draft (G stands for Galitz) was designed following the guidelines from Galitz (2007). Data was still organized in a hierarchical tree at the left of the interface (see Figure 2). Galitz (2007) suggests the use of a hierarchical menu structure when many relationships between the selectable alternatives exist. The difference to the E-Draft was the tree length and the level of detail. In this draft, the 7 +/- 2 rule (Miller, 1956, cited in: Dix et al., 2004) was applied to achieve a more memorable and less complex menu structure. This was enabled by adding functionality to the map. When clicking on a gauging station on the map, a selection list of available information of the selected place appeared. This allowed shifting the spatial navigation from the tree to the map and therefore minimizing the length of the tree. Another difference to the E-Draft was the absence of tabs. Every information request caused the generation of a new window, which was loaded underneath the active one at the bottom of the map resulting in a stack of windows holding distinct information. This GUI was meant to present a mixture between the E-Draft's non-geographical interface and an information system with a clearly spatial navigation ability.

Figure 2. The G-Draft based on guidelines by Galitz 2007

*The C-Draft*

The C-Draft's design (C stands for Cooper) followed the guidelines from Cooper et al. (2007). The most obvious change was the absence of the navigation tree on the left-hand side of the interface (see Figure 3). Instead, a large button with a filing cabinet symbol allowed the user to navigate thematically. This change was based on Cooper et al.'s (2007) opinion that hierarchic menu structures are not ideal for depicting small quantities of categories. They suggest using monocline groupings, which never exceed one level of nesting. By clicking on the filing cabinet button, a simple and sorted list with all available information appeared. Due to the missing thematic navigation bar on the left, the information area was very large compared to the other drafts and filled to the major extent by the map. It was therefore possible to display several types of gauging stations simultaneously on one map. Information about a specific gauging station could be obtained by clicking on the station. This opened a context menu which allowed the user to choose between the available parameters at the selected station. Information about a specific station was presented in a pop up window above the base map. Unlike the GUIs introduced above, in the C-Draft new information was loaded on top of the previous visualization. Scrolling was therefore not required. The C-Draft was designed to represent an information system with a spatial navigation focus to the greatest possible extent.

Figure 3. The C-Draft based on guidelines by Cooper et al. 2007

*The user experiment*

In a second step, the design drafts served as visual stimuli in an experiment with test subjects combining classical usability evaluation methods and eye movement recording. 21 (6 female, 15 male) representative test subjects have been selected based on their experience, training and familiarity with similar systems. The average age of the participants was 32 with a standard deviation (SD) of 8.5 years.

At the beginning of the recording session, the participants were informed about the procedure of the experiment. The eye tracker was then calibrated and the the first stimulus was presented to the participant. The experiment leader then explained the first working task verbally. The tasks for the experiment were designed to be realistic and reflecting typical tasks to be solved with the platform, but simple enough to avoid misunderstandings. The answers were given verbally as well and were noted by the experiment leader, hence the contact between the participants and the eye tracker was maintained throughout the recording. A within-subjects design was chosen to obtain the most information from the experiment. Each participant was therefore presented with all three stimuli, i.e. the three drafts. As a consequence of this, the sequence of the stimuli presented to the participants was counterbalanced in a controlled way using a Latin Square. The number of 21 participants was optimal for the chosen design. As the comparison of the tasks was not a relevant factor, the sequence of tasks was held constant for each participant.

After all four working tasks had been completed for one stimulus, the recording was interrupted. The participants were then asked to fill in a standardized System Usability Scale questionnaire (SUS; Brooke, 1996) to rate their satisfaction with the interface. SUS is a usability measurement tool that produces a usability scale for a system, ranging from 0 to 100, 100 being the best possible score. It therefore delivers a scale that enables comparison of systems. The SUS questionnaire is composed of ten statements that have to be rated by the user from one to five on a Likert Scale. A systematic review of five usability questionnaires showed that SUS yielded the most reliable results across sample sizes, even though it was the simplest questionnaire studied (Tullis and Stetson, 2004). Additionally, Bangor et al. (2008) point out that SUS provides a good and valid method of comparing different interfaces' usability.

The overall duration of the experiment was approximately 50 minutes. The test persons were offered no compensation for the participation in the experiment.

*Aparatus*

The experiment was run on a PC running Windows XP Professional. The stimuli were displayed on a 20 inch Flat Panel (LCD) with 1280 x 1024 pixels resolution in Microsoft Internet Explorer 7 in full screen mode. During the experiment, video and sound were recorded by a web camera in order to observe the participant during the experiment. Tobii X120, a video-based combined pupil and corneal reflection table mounted binocular eye tracker was used to record the gazes of participants with a 60 Hz sampling resolution.

## Results

The results from the SUS questionnaire and the eye movement recordings are mostly presented as mean values and standard deviations (SDs). If the data was ordinal scaled or the SDs were high, the median (M) was calculated. The Kolmogorov-Smirnov-Test was applied to every interval scaled variable in order to identify if the data was normally distributed. Normally distributed data was compared through use of either the T-Test for two independent samples or the one-way analysis of variance (ANOVA) combined with the Duncan Test for homogenous subgroups for more than two independent samples. Not normally distributed data was compared using the Mann-Whitney U Test (for two independent samples) and the Kruskal-Wallis H Test (for more than two independent samples). It was found that all eye tracking metrics, including the time per task and the number of mouse clicks, showed a significant deviation from the normal distribution. A 5% level of significance was used for all statistical tests.

*Results from classical usability evaluation methods*

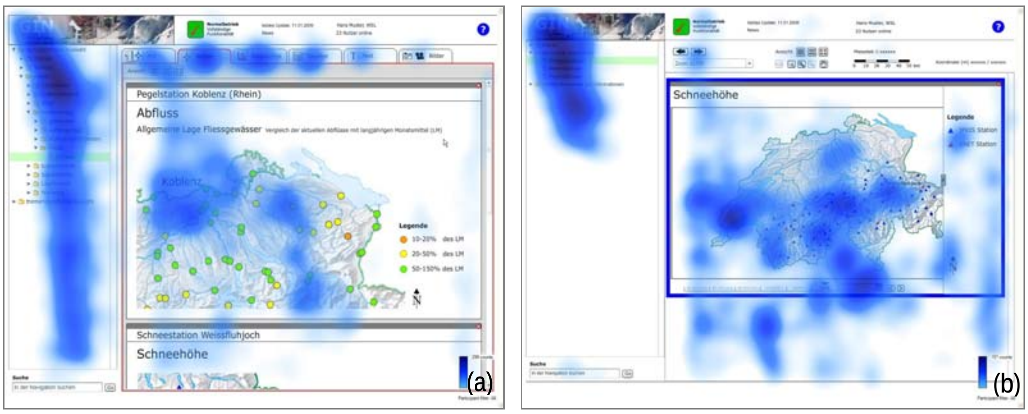
In the SUS, the G-Draft scored best with an average of 81.55 and a SD of 12.03. The C-Draft reached a SUS score of 60.00 (SD: 21.67) and the E-Draft obtained 56.55 (SD: 16.91). The SUS scores are significantly different ($F = 12.848$, $p < 0.001$) among the stimuli. The Duncan Test revealed that the G-Draft's SUS scores are significantly higher than the E- and C-Draft's ratings. The difference between the latter two GUIs however was not significant. Furthermore, they make statements about acceptable SUS scores and interpret them. According to (Bangor et al., 2008, p. 592), the E-Draft can be classified as “*OK*”, the C-Draft between “*OK*” and “*Good*” and the G-Draft between “*Good*” and “*Excellent*”. In the same breath, they point out that this empirical classification is not generally valid but can help ranging the rated systems in a wider and absolute context of usability.

In terms of the average time per task taken by all 21 participants, the G-Draft performed best (M: 24.04s, SD: 26.94), the C-Draft intermediate (M: 30.29s, SD: 56.44) and the E-Draft worst (M: 65.31, SD: 61.56). The means of the total times used for the different stimuli differ significantly ($p < 0.001$). Specifically, the E-Draft’s task completion times are significantly different to those of the G-Draft ($Z = -7.604$, $p < 0.001$) and those of the C-Draft ($Z = -5.114$, $p < 0.001$). The G-Draft’s task completion times however are not significantly different to the C-Draft’s performance ($Z = -1.927$, $p = 0.054$) too. The best result of the G-Draft in terms of efficiency is also evidenced by the smallest SD of the total mean durations of the stimuli.

In addition, the number of mouse clicks needed to complete all tasks on a stimulus differs significantly among each of the three stimuli ($p < 0.001$). More specifically, the E-Draft’s number of mouse clicks (26, SD: 19) is significantly higher than the mouse clicks needed for the G-Draft (12, SD: 9, $Z = -7.148$, $p < 0.001$). The E-Draft also required a significantly higher number of mouse clicks compared to the C-Draft (17, SD: 14, $Z = -4.844$, $p < 0.001$). Furthermore, in terms of the number of mouse clicks, the G-Draft performed significantly better than the C-Draft ($Z = -2.221$, $p = 0.026$).

*Results from the eye tracking evaluation*

The quality of the eye tracking data was not equally high for every participant. Due to failures in several recordings, three records for the E- and C-Draft as well as two records for the G-Draft had been excluded from the dataset used for analysis.
For every stimulus, a fixation density map was created (see Figure 2). Fixation density maps visualize all fixations that occurred on a stimulus during an experiment. They consist of a static picture of the stimulus, overlaid with an interpolated fixation density surface. Darker areas specify regions with more fixations.



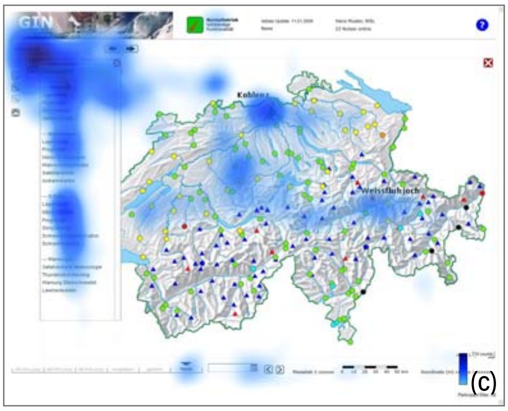



Figure 2. Fixation density maps of (a) the E-Draft, (b) the G-Draft and (c) the C-Draft

The spatial dispersion of the fixations on the interfaces reveals a clustered pattern. For this reason the GUIs have been divided into areas of interest (AOIs). Table 1 provides a selective overview on the most important AOIs and the associated eye tracking results.

| Stimulus | AOI Name | Percentage of screen [%] | Total no. of fixations | Median no. of fixations per task | Mean Fixation duration [ms] |
|---|---|---|---|---|---|
| E-Draft | Navigation bar | 17.58 | 5317 | 37.5 | 523.50 |
| | Tabs section | 2.90 | 773 | 6.5 | 398.25 |
| | Information window | 57.29 | 5566 | 54.5 | 350.47 |
| G-Draft | Navigation bar | 21.65 | 1047 | 8.5 | 461.77 |
| | Toolbar | 5.12 | 147 | 0 | 291.78 |
| | Information window | 53.20 | 4079 | 32.5 | 380.56 |
| C-Draft | Navigation window | 10.61 | 2623 | 22.5 | 515.76 |
| | Filing cabinet button | 0.10 | 90 | 0 | 453.02 |
| | Information window | 68.62 | 5813 | 53 | 425.08 |

Table 1. AOIs and dedicated eye tracking results

The complex tree-navigation in the E-Draft generated the most and the longest fixations. The G-Draft produced significantly less fixations than the E-Draft ($Z = -8.581$, $p < 0.001$). The number of fixations on the C-Draft's navigation window could not be analyzed because it was only visible if the filing cabinet button was active. The number of fixations on the navigation window is therefore incorrect due to its dynamic character. It rather describes the fixations that occurred in the area where the navigation window appeared if the cabinet button was clicked. Furthermore, the fixations in the dynamic information windows of all three drafts are not directly referable to the content depicted on the static background layers in Figure 2 due to the fact that the participants had to interact with the dynamic content to find the required information. On the other hand, the number of fixations and the fixation durations reported in Table 1 are still meaningful because these numbers regard the overall performance of the navigation windows themselves. The difference between the fixation durations on the E-Draft's and the G-Draft's navigation bar is significant ($Z = -9.261$, $p < 0.001$). It can therefore be assumed that the navigation bar in the G-Draft was less complicated to process mentally, than in the E-Draft. The fixation durations on the C-Draft's navigation window are not meaningful because the window was only visible if the filing cabinet button was active. The tabs section in the E-Draft generated relatively small mean fixation durations. It can therefore be concluded that the tab function was clear and easily understood. The filing cabinet button on the C-Draft, which was essential to solve all four tasks, did not generate long fixation durations as well.

Overall, the G-Draft was observed to be the most usable GUI draft. In both classical measures of usability and eye tracking analysis it achieved the best results most often. It is likely that the good SUS rating, the short task completion times and the low number of mouse clicks can be explained by the G-Draft's short and low level navigation tree. Galitz (2007) suggests the use of a well-balanced and where possible two-leveled hierarchical navigation tree to organize a system's content. The G-Draft's navigation tree therefore uses mostly two, maximum three levels. The significantly shorter length of time needed to complete the tasks in the G-Draft compared to the E-Draft supports the Hick-Hyman Law (Hick, 1952, Hyman, 1953, cited in: Cockburn et al., 2007), which claims that the time taken to choose an item (for example on an interface) is proportional to the total number of selectable items. The eye tracking data also backs up these results. The number of fixations and the average fixation durations in the G-Draft's navigation section are both significantly lower than for the E-Draft. Due to the fact that both the G-Draft and the E-Draft are similar in design (the biggest difference is the complexity of the tree) and the fact that the E-Draft's navigation to the gauging stations had to be done inside the tree, it can be assumed that spatial navigation is more efficiently performed on a map.

In general, the G-Draft produced the lowest mean and median number of fixations, as well as the shortest average fixation duration. This supports the idea that a combination of the classical and well established tree navigation, combined with interactive map functions, is helpful when designing an efficient user interface. The C-Draft, as the representative of an interface employing a spatial navigation strategy, performed second best in terms of usability. Due to the facts explained above, the E-Draft can be regarded as the GUI with the poorest usability among the three stimuli.

## Conclusions

The results of this study suggest that the existing visualization theories and design guidelines, developed for non-spatial interactive systems, can be partially applied to the spatial domain. In particular, it was shown that tasks which are not related to spatial objectives, such as thematic navigation, should be designed according to the classic theories. It was revealed that the users like metaphors they already know and that systems employing them are more user friendly.

Yet, an independent theory to support methods for usable geovisualization is regarded as important. However, GIScience should not reinvent the wheel. According to the results of this study, the G-Draft represents a possible visualization concept for a usable interactive spatial information system. It could be complemented with tabs to sort the different data visualizations.

In general, it is however not possible to make universally valid statements with the results of only one study. Further research in geographic information science (GIScience) is therefore required in order to deliver useful and empirically proven design guidelines for usable geographic information interfaces. Furthermore, usability evaluations of basic map functions such as zooming, panning, rotating or tilting for 3D solutions should be targeted. A framework of comparable and complementary studies could help to extend the existing guidelines for usable non-spatial computer interfaces. Some work has already been done (e.g. by Nivala et al., 2008, Harrower and Sheesley, 2005, Hornbæk et al., 2002, You et al., 2007) which shows that GIScience is progressing in this field. Additionally, eye tracking is regarded as a promising technique for enhancing the knowledge of the effectiveness of interactive geovisualization tools. In doing so, fundamental experiments are required to enable a bottom-up approach.

**References**


BANGOR, A., KORTUM, P. T. & MILLER, J. T. (2008) An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction,* 24**,** 574-594.

BROOKE, J. (1996) SUS - A quick and dirty usability scale. IN JORDAN, P. W., THOMAS, B., WEERDMEESTER, B. A. & MCCLELLAND, A. L. (Eds.) *Usability Evaluation in Industry.* London, Taylor and Francis.

COCKBURN, A., GUTWIN, C. & GREENBERG, S. (2007) A predictive model of menu performance. *Conference on Human Factors in Computing Systems CHI.* San Jose, CA, ACM.

ÇÖLTEKIN, A., HEIL, B., GARLANDINI, S. & FABRIKANT, S. I. (2009) Evaluating the Effectiveness of Interactive Map Interface Designs: A Case Study Integrating Usability Metrics with Eye-Movement Analysis. *Cartography and Geographic Information Science,* 36**,** 5-17.

COOPER, A., REIMANN, R. & CRONIN, D. (2007) *About face 3: the essentials of interaction design,* Indianapolis, IN, Wiley Pub.

DIX, A., FINLAY, J., ABOWD, G., D & BEALE, R. (2004) *Human-computer interaction,* Harlow, Pearson/Prentice-Hall.

DUCHOWSKI, A. T. (2007) *Eye tracking methodology: theory and practice,* London, Springer.

FABRIKANT, S. I. (2005) Towards an understanding of geovisualization with dynamic displays. Issues and Prospects. *Proceedings, American Association for Artificial Intelligence (AAAI) 2005 Spring Symposium.* Stanford, CA.
FUHRMANN, S., AHONEN-RAINIO, P., EDSALL, R. M., FABRIKANT, S. I., KOUA, E. L., TOBÓN, C., WARE, C. & WILSON, S. (2005) Making Useful and Useable Geovisualization: Design and Evaluation Issues. IN DYKES, J., MACEACHREN, A. M. & KRAAK, M.-J. (Eds.) *Exploring Geovisualization.* Oxford, Elsevier.
GALITZ, W. O. (2007) *The essential guide to user interface design: an introduction to GUI design principles and techniques,* Indianapolis, IN, Wiley Pub.
HARROWER, M. & SHEESLEY, B. (2005) Designing Better Map Interfaces: A Framework for Panning and Zooming. *Transactions in GIS,* 9**,** 77-89.
HICK, W. (1952) On the rate of gain of information. *Journal of Experimental Psychology,* 4**,** 11-26.
HORNBÆK, K., BEDERSON, B. B. & PLAISANT, C. (2002) Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM Transactions on Computer-Human Interaction (TOCHI),* 9**,** 362-389.
HYMAN, R. (1953) Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology,* 45**,** 188-196.
MACEACHREN, A. M. & KRAAK, M.-J. (2001) Research Challenges in Geovisualization. *Cartography and Geographic Information Science,* 28**,** 3-12.
MILLER, G. A. (1956) The magical number seven, plus minus two: Some limits on our capacity for processing information. *Psychological Review,* 63**,** 81-97.
NIVALA, A.-M., BREWSTER, S. & SARJAKOSKI, T. L. (2008) Usability Evaluation of Web Mapping Sites. *The Cartographic Journal,* 45**,** 129-138.
SLOCUM, T. A., BLOK, C., JIANG, B., KOUSSOULAKOU, A., MONTELLO, D. R., FUHRMANN, S. & HEDLEY, N. R. (2001) Cognitive and Usability Issues in Geovisualization. *Cartography and Geographic Information Science,* 28**,** 61-75.
TULLIS, T. S. & STETSON, J. N. (2004) A Comparison of Questionnaires for Assessing Website Usability. *Usability Professionals Association Conference UPA.* Minneapolis, MN.
YOU, M., CHEN, C.-W., LIU, H. & LIN, H. (2007) A Usability Evaluation of Web Map Zoom and Pan Functions. *International Journal of Design,* 1**,** 1-11 Online journal, http://www.ijdesign.org/ojs/index.php/IJDesign/article/viewFile/31/4 Accessed 2009-05-22.