Topographic Maps Based on Kohonen Self Organizing Maps
An Empirical Approach

Thomas Mandl*, Maximilian Eibl**

*University of Hildesheim, Information Science, IFAS
Marienburger Platz 22, D-31141 Hildesheim, Germany
Phone: ++49-5121-88 38 37, Fax: ++49-5121-88 38 02
email: mandl@uni-hildesheim.de

** GESIS
Schiffbauerdamm 19, D-10117 Berlin, Germany
Phone: ++49-30-30874-271, Fax: ++49-30-28 23 692
email: eibl@berlin.iz-soz.de

ABSTRACT: Two-dimensional maps are a valuable interface element for the visualization of information retrieval results or other large sets of objects. Various methods exist for the creation of these maps. This article describes a comparative evaluation of topographic maps based on “Kohonen Self Organizing Maps” (SOM). These results show that the mapping method has to be chosen very carefully and different methods should be tested for an application.

KEYWORDS: self organizing maps, Kohonen maps, topic maps, 2D maps, evaluation, information retrieval, information visualization

1 INTRODUCTION

Apart from keyword search, browsing is the second most important strategy in information seeking: „browsing is an approach to information seeking that is informal and opportunistic and depends heavily on the information environment“ ([13]:100). Even though, browsing is a rather informal approach is sometimes unpredictable, evaluations of browsing interfaces are necessary.
Several approaches have been built which aim at facilitate browsing in document sets. One of those approaches uses visualization based on neural reduction algorithms. Neural networks are powerful computing algorithms on human information processing which have been successfully applied in information retrieval [10] and digital libraries [6].
which are able to intelligently map objects into a topographic map. Their capabilities can be used to construct two dimensional graphic displays in order to support browsing strategies in a graphic environment.
Topographic maps can be regarded as multimodal interfaces taking advantage of graphical and language elements in which the positioning and the text of an object jointly express the meaning of an object..

2 TWO DIMENSIONAL TOPOGRAPHIC DOCUMENT DISPLAYS

Two dimensional document maps have received considerably attention in the last few years. Several experimental and commercial systems have been developed. Examples that are available via Internet are Cartia (http://www.cartia.com) or ET-Map (http://ai2.bpa.arizona.edu/ent/entertain1).
2D maps are an important tool to enhance human computer interaction with a growing number of applications. Most often, text documents are arranged in two-dimensional maps, however, other information objects are also being ordered. Merkl applied Kohonen networks to software code in order to improve reusability [14]. Rauber and Frühwirth created a SOM of music pieces [16]. Coleman stresses the importance of these systems within educational software [3]. They point out, that demonstrating and visualizing the interrelationships between concepts is an important objective in teaching. Mutschke used social network theory to compute centrality of authors in scientific discourse. These communities are subsequently visualized to allow path analysis and advanced retrieval techniques [15]. News articles

are subject to two dimensional display in Newsmaps (http://www.newsmaps.com). This variety of systems in the electronic service area demonstrates the attraction of these methods to researchers. Further applications in finance and trade are likely to be developed.
These maps visualize a document set which is usually reduced by a prior information retrieval query. They are based on the metaphor of maps or landscapes. The use of this metaphor is of varying strength. Whereas the maps in Cartia really look like topographical maps, ET-Maps produces rather abstract versions. An example for such a map is shown in Figure 1:

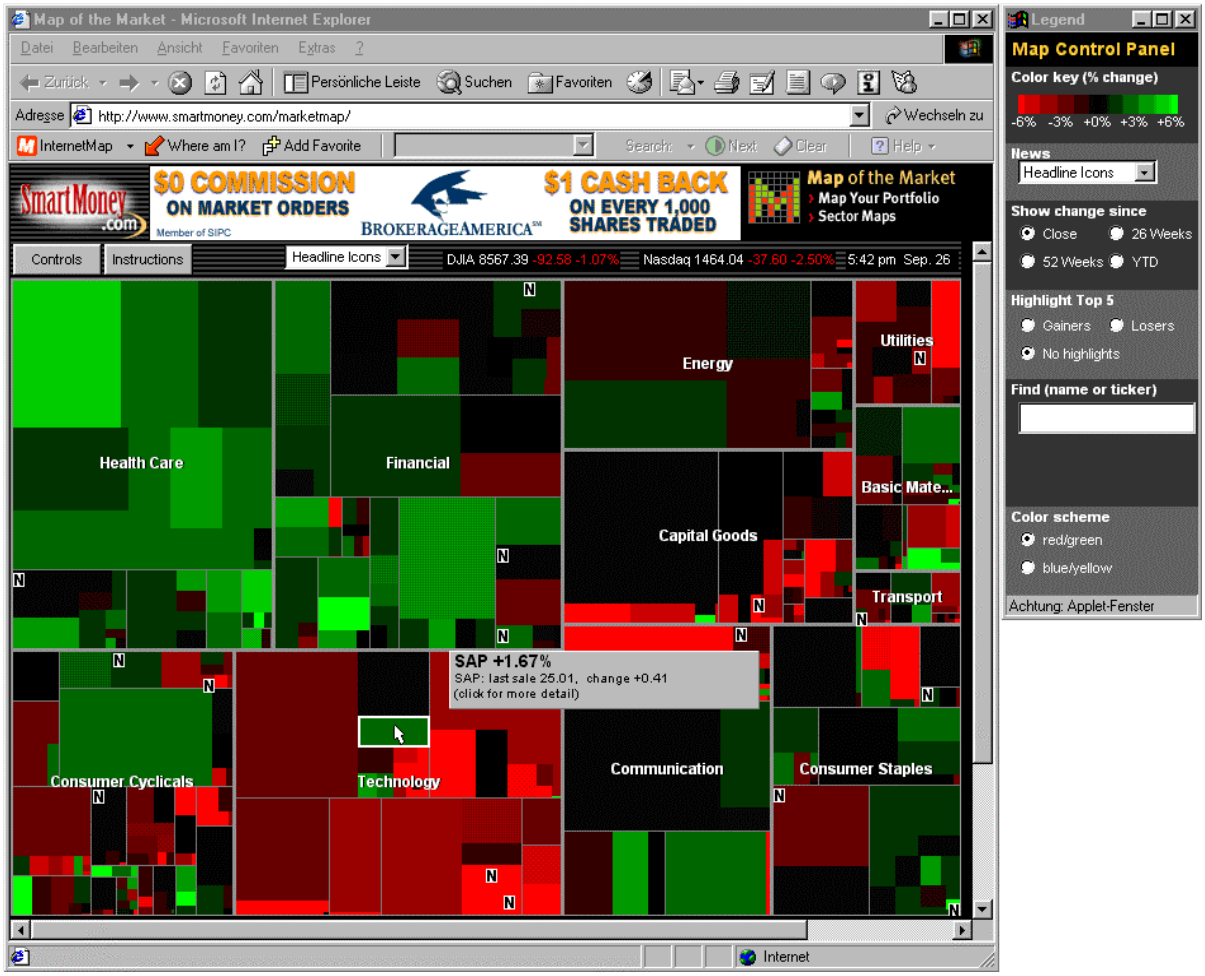

Figure 1: Map of the performance of more than 600 stocks by http://www.smartmoney.com. A green field in the "Technology"-area is highlighted by moving the mouse over this field. A popup window gives detailed information about this specific stock.

Figure 2 demonstrates a schematic map to a data set containing documents about "visualization". This map is a very simple one. It constructed three elevations which refer to the three mayor topics in this document set: "information visualization", "scientific visualization", and "computer graphics". Each elevation symbolizes a collection of documents relevant to the mentioned topic. The elevations are surrounded by the sea containing no documents.

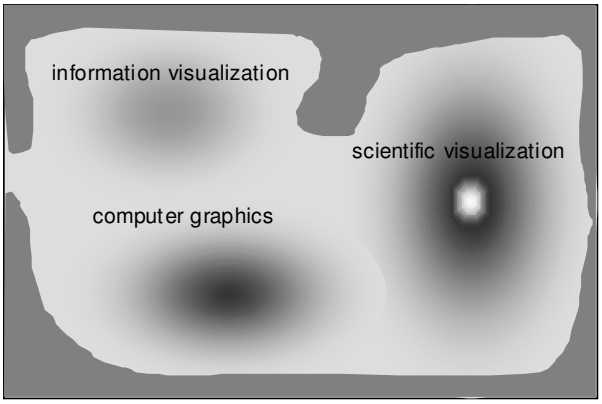

Figure 2: Topographic Map for "visualization": The dark colored regions symbolize aggregations of documents

In this little example the elevations are constructed manually and therefore very simple and comprehensible. In reality those maps are self organizing. The resulting document elevations and the final appearance of the map can not be foreseen since the resulting topics are identified by the system using strictly statistical methods. An elaborated semantic approach is not yet available.

The unpredictability of this approach is an advantage and a disadvantage at the same time. On the one hand this approach facilitates the recognition of document clusters. Those clusters sometimes describe new and yet unknown topics. On the other hand these clusters are frequently completely useless since semantic aspects are omitted. Here, the introduction of XML could lead to an improvement.

Nevertheless, maps are one of the design strategies that will be of great importance for the future construction of search engines. Document maps try to exploit the visual capabilities of humans in order to create interfaces easy to use. Objects closely related are located next to each other, geometric distance becomes a metaphor for semantic similarity. This design strategy is cognitive plausible and is usually well understood by users.

Finding information in databases can be realized by explicit search or by browsing. Topographical maps are frequently based on a broad database search. But concerning the user interaction they emphasize the browsing aspect of information gathering.

The main technical aspects of creating two-dimensional maps lays in the reduction of data dimensions. Document collections the A popular method for creating such maps are Kohonen Self-Organizing Maps (SOM). A SOM is a clustering algorithm which guarantees, that clusters close to each other are similar [1]. Other methods are latent semantic indexing (LSI, [1]) or factor analysis.

3 EVALUATION OF BROWSING INTERFACES

As noted in [3] and [11] there is a lack of evaluation concerning visualizations in general. A a methodology to compare the different methods for designing two-dimensional map layouts are proposed in [11]. The relationships between all objects of the resulting display are compared. The following pseudo-code demonstrates the algorithm:

```
For all documents di
     For all documents dj
          Compute Similarity di, dj using method m1
          Compute Similarity di, dj using method m2
     Build ranked list for di, dj using m1
     Build ranked list for di, dj using m2
     Compare lists using the Spearman correlation coefficient
Compute average correlation coefficient
```

The central measure used in this comparison is the Spearman correlation coefficient [5] which has already been applied for the evaluation of IR systems in order to measure the differences between similarity matrices derived by different algorithms [12]. The evaluation showed that the resulting document distribution differed largely one from another:

Table 1: Correlation

Number of documents	1000	100 (average of 10 sets)	10 (average of 100 sets)	Highest correlation for a set of 10 documents
Correlation	0,053	0,037	0,013	0,34

4 FURTHER EVALUATION OF BROWSING INTERFACES

The evaluation method presented in [11] , a document is chosen as starting point and the user is assumed to assess the other documents in order of their distance from this starting document. However, considering typical browsing behaviour, it can be assumed, that users have a tendency to assess documents in the reading direction. Therefore, documents below and at the right of the starting point are more likely to be considered consequently.
The new evaluation method presented here takes this into account. Only the n closest documents which are below or at the right side of the starting document are considered. In this case, n must be chosen rather low. We experimented with ten and twenty documents. Obviously, the orientation of the map has to be considered as well now. Once a Kohonen SOM has been calculated, it can be rotated at an angle of 90° without changing the mapping from the input space into the Kohonen space:

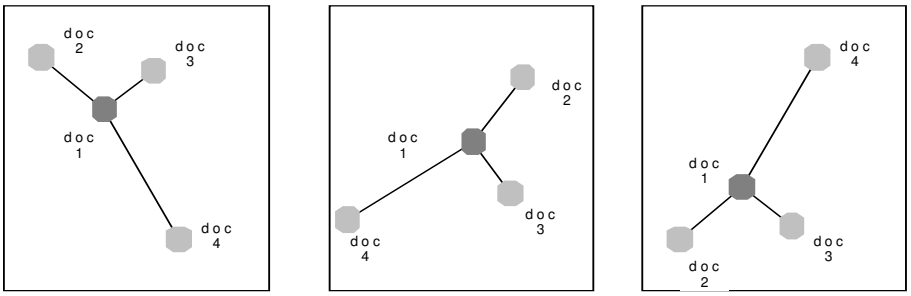

Figure 3: Variation of the display

The evaluation method considers typical user behavior in browsing interfaces in general and in two-dimensional maps especially. Starting point is one interesting document, which the user has identifies by unguided browsing or through keyword search. Systems combining keyword search and browsing in maps are common.
A naive question could be the following: are topographic maps created with different methods merely variations created by rotations? As could be shown, topographic maps differ greatly according to the distance between objects [11]. Therefore, different maps can not be simple identical copies which have been rotated. However, considering the actual user behavior, the maps could be much more similar. In that case, the rotation could be more important than the creation algorithm.
At the moment the following evaluation methods are tested:

- Orientation correction
- Chain tracking: Users are expeced to find one interesting document and assess the document closest to it next and so forth.
- Density evaluation: The visual attention of users is expeced to be attracted by very central and dense groups of objects first.

First tests seem to confirm the results in described [3] . Nevertheless, further experiments need to be conducted. Another important aspect which needs to be evaluated intensively is filtering as regard to adaptive hypertext systems. Users do not need to see all objects in all cases and personalized overview maps are one important goal of user models [7, 8]. Once a user model has been developed, the best method for each user may be mapped to the users parameters.

5 CONCLUSION

The evaluation of topographic maps is extremely important. There is a growing number of systems available and little is known about their usability. Few experiments have been carried out, especially with regard towards the best algorithms. This article extends a evaluation methodology which does not require expensive user tests and leads to quantitative results.

REFERENCES

[1] Berry, M., Dumais, S.; Letsche, T. (1995): Computational Methods for Intelligent Information Access. In: Proc. ACM Supercomputing '95. San Diego, CA. pp. 1-38.

[2] Bose, N.; Liang, P. (1996): Neural Network Fundamentals with Graphs, Algorithms, and Applications. New York et al.

[3] Coleman, A.; Smith, T.; Buchel, O.; Mayer, R. (2001): Learning Spaces in Digital Libraries. In: Constantopoulos, P.; Solvberg, I. (eds.): 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001) Darmstadt Sept. 4.-8. Berlin et al.: Springer. pp. 251-262.

[4] Hammwöhner, R. (1997); Offene Hypertextsysteme – das Konstanzer Hypertextsystem (KHS) im wissenschaftlichen und technischen Kontext; Konstanz: Univ.-Verl.

[5] Hartung, J. (1984): Lehr- und Handbuch der angewandten Statistik. München, Wien.

[6] Hellweg, H.; Krause, J.; Mandl, T.; Marx, J.; Müller, M; Mutschke, P.; Strötgen, R. (2001): Treatment of Semantic Heterogeneity in Information Retrieval. IZ-Arbeitsbericht Nr. 23, IZ Sozialwissenschaften, Bonn. http://www.gesis.org/Publikationen/Berichte/IZ_Arbeitsberichte/index.htm#ab23

[7] Kobsa, A.; Koenemann, J.; Pohl, W. (2000); Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships; St. Augustin: GMD FIT; http://fit.gmd.de/~kobsa/papers/Kobsa-PHPT-draft.doc

[8] Kobsa, A.; Specht, M. (1999); Interaction of domain expertise and interface design in adaptive educational hypermedia; Universität Konstanz; http://fit.gmd.de/~kobsa/papers/1999-WWW8UM99-kobsa.pdf

[9] Eibl, M. (2000). Visualisierung im Document Retrieval. Theoretische und praktische Zusammenführung von Softwareergonomie und Graphik Design. IZ-Forschungsbericht 3, Informationszentrum Sozialwissenschaften, Bonn. (includes CD-ROM with English summary and animation)

[10] Mandl, T. (2000): Tolerant Information Retrieval with Backpropagation Networks. In: Neural Computing & Applications. Special Issue on Neural Computing in Human-Computer Interaction. Vol. 9 (4). pp. 280-289.

[11] Mandl, T.; Eibl, M. (2001): Evaluating Visualizations: A Method for Comparing 2D Maps. In: Proceedings of the HCI International 2001 (9th International Conference on Human-Computer Interaction), New Orleans, August 2001. Mahwah, NJ; London: Lawrence Erlbaum Associates. pp. 1145-1149.

[12] Mandl, T. (1998): Learning Similarity Functions in Information Retrieval. In: Zimmermann, H.J. (ed.): EUFIT '98. 6th European Congress on Intelligent Techniques and Soft Computing. Aachen, Germany, 8.-10.9.1998. pp. 771-775.

[13] Marchionini, G. (1995): Information Seeking in Electronic Environments.

[14] Merkl, D.; Tjoa, AM.; Kappel, G. (1994): Learning the Semantic Similarity of Reusable Software Components. In: Frakes, William B. (eds.): Proceedings of the Third International Conference on Software Reuse: Advances in Software Reusability. Rio de Janeiro, 1.-4.11. 1994. Washington et al. pp. 33-41.

[15] Mutschke, P. (2001): Enhancing Information Retrieval in Federated Bibliographic Data Sources Using Author Network Based Stratagems. In: Constantopoulos, P.; Solvberg, I. (eds.): 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001) Darmstadt Sept. 4.-8. Berlin et al.: Springer. pp. 287-299.

[16] Rauber, A.; Frühwirth, M. (2001): Automatically Analyzing and Organizing Music Archives. In: Constantopoulos, P.; Solvberg, I. (eds.): 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001) Darmstadt Sept. 4.-8. Berlin et al.: Springer. pp. 402-414