INCLUDING USER STRATEGIES IN THE EVALUATION OF GRAPHIC DESIGN INTERFACES FOR BROWSING DOCUMENTS

Maximilian Eibl
Social Science Information Centre
Schiffbauerdamm 19, 10117 Berlin, Germany
eibl@berlin.iz-soz.de
http://www.gesis.org

Thomas Mandl
Information Science, University of Hildesheim, Marienburger Platz 22, Germany
mandl@uni-hildesheim.de
http://www.uni-hildesheim.de/~mandl

ABSTRACT

This article presents an analytical evaluation method for innovative and graphically rich displays beyond the common WIMP interfaces. This method is based on user strategies and compares different algorithms for creating 2D or 3D environments and leads to a correlation measure which can be consulted in the creation process of such environments. The results suggest, that the graphical algorithm may have more influence on the quality of maps than user strategies. Therefore, the correct choice of appropriate arithmetic algorithms is crucial. Furthermore, the article discusses the extension of the evaluation method to 3D environments for semantic organization of homogeneous objects. By presenting this method and its results, an example is given for rigid and user-oriented evaluation of advanced graphical interfaces which are applied in the internet.

Keywords: topographic map, 2D maps, evaluation, visualization, human computer interaction, browsing interfaces, user strategies

1. INTRODUCTION

Computer Graphic algorithms lead to new opportunities for user interfaces. Unlike common WIMP (windows, icons, menus, pointers) interfaces, advanced graphic elements consist of more than lines and rectangles and partly rely on the aesthetic perception by the user. Vague human needs are addressed through advanced computer graphic routines.

Concerning the evaluation of such interfaces, the scientific community of human-computer interaction is far behind the technological development. Little is actually known about the perception of aesthetic aspects or the influence of strongly graphical presentations. Evaluation is nevertheless extremely important, as developers are confronted with a significant number of design decisions and often need to choose among several competing algorithms.

An evaluation procedure needs to include vague human factors as well as computational aspects of computer graphics. The prohibitive high costs for intensive user experiments call for additional analytic evaluation methods which allow the rigid assessment of at least some aspects of advanced human-computer interfaces. In web design, heuristic evaluation serves this purpose. This paper also presents a relatively cheap evaluation methodology.

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“ [March95:100]. Since browsing is a rather informal approach, it is sometimes quite unpredictable. This

makes the evaluation of browsing interfaces even more necessary.

Typical sites for browsing, are internet catalogs like Yahoo![1]. Unlike these text oriented systems, many sites turn to graphical presentations in order to overcome the limitations of screen space. Graphical representations use the given space far better than textual ones. While some systems like WebBrain[2] rely only partially on these features, others like WebMap[3] are primarily graphically orientated.

In order to support browsing following user needs, objects are displayed in a way such that the semantic relationships between them can be visually perceived. The perception of many details can be fostered by using a graphical display instead of presenting merely numbers or tables. However, even powerful computer graphic systems often cannot present the whole complexity of a problem.

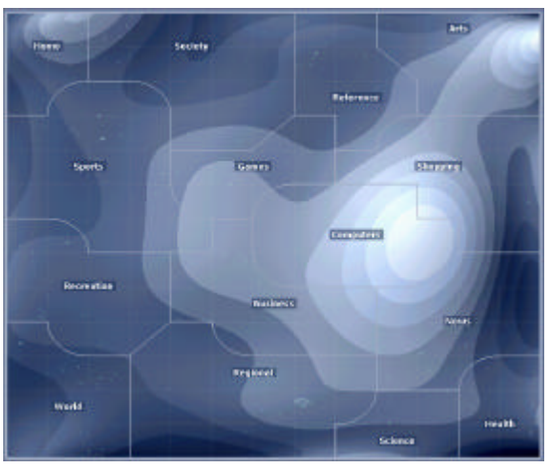

Figure 1: A topographic map
http://www.webmap.com

2. TWO-DIMENSIONAL TOPOGRAPHIC OBJECT DISPLAYS

Two-dimensional document maps have received considerable attention in recent years. Several experimental and commercial systems have been developed[4]. Certainly, two-dimensional 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 which are easy to use. Objects closely related are located next to each other, geometric distance becomes a metaphor for semantic similarity. This design strategy is cognitively plausible and is usually well understood by users. This metaphor is one of the most typical metaphors used in graphical interfaces and is close to being a visual formalism. It is discussed in the context of metaphoric interfaces and visual formalisms in [Eib01].

The basic spatial metaphor is usually enriched with other design elements like color, lines as cluster boundaries, three-dimensional effects and text fields. Some of these elements can be seen in figure 1.

In experimental systems, 2D maps have been applied to a variety of different object types including software code [Merkl94] [Ye00], economic time series [Taska01], authors of scientific literature and their position in a scionometric network of discourse [Mutsc01], newsgroup postings [Oja99], music pieces [Raube01] and multimedia internet content [Rouss99]. These maps seem to be especially suited for image retrieval. In this vague domain which requires the visual perception of the objects, the display of their similarity is quite a natural way of access [Koske01], [Ojala01]. Two-dimensional maps are often referred to as data mining tools since they reveal complex relationships between objects and clusters of objects and allow humans to visually perceive this knowledge naturally and easier than from e.g. numeric tables.

A popular method for visual formalism is the so called Kohonen Self-Organizing Map (SOM). A SOM is a clustering algorithm which guarantees, that clusters close to each other contain similar items [Bose96]. Further possibilities to create two dimensional maps lie in using other dimensionality reduction algorithms like singular value decomposition or factor analysis.

3. EVALUATION METHOD

Though many approaches to information retrieval incorporate visualization techniques only few evaluations have been conducted. Appropriate quantitative evaluation methods for evaluations have not been established yet. Most formal studies rely on the standard information retrieval measures recall and precision.

[Eibl00] reports on 25 visualizations for information retrieval but only two formal evaluation studies aiming at quantitative data could be found: SENTINEL [Knepp98] and J24 [Odgen98]. Further user tests were conducted by [Swan98] and [Chen96] who conducted a user test for a SOM. In this study, the qualitative results were promising. However, only subjective

1 http://www.yahoo.com
2 http://www.WebBrain.com
3 http://www.WebMap.com
4 e.g. http://www.cartia.com, http://websom.hut.fi/websom

evaluation methods were applied in which the test users showed a positive attitude towards the maps. No comparison to other search methods was carried out in order to prove the superiority of the maps, nor was the SOM compared to other algorithms to create 2D maps.

A method for assessing 2D maps has been suggested in [Mandl01]. However, it does not take different user strategies into account but merely compares the correlation between algorithms assuming homogeneous user behavior. In this article, the method is extended to take both a local and a global browsing strategy into account. Experiments with a real world document corpus suggest that the correlation between maps created by different methods is lower than the correlation between different user strategies.

3.1 BASIC EVALUATION METHODOLOGY

The primary goal of the evaluation presented in [Mandl01] was to find a rationale for the use of certain mapping methods into the two-dimensional space. Usually, no reasons are given for the choice of a specific algorithm like Kohonen SOM. For that reason, the first question which needed to be asked was whether these methods lead to different maps at all or whether these differences are negligible. The difference needs to be assessed from the users' point of view.

Usually, the formal evaluation of information retrieval systems requires a ranked list of objects which is not provided by the two-dimensional map. Therefore, the two-dimensional map is transformed into a ranked list which represents the order in which the user browses through the documents. The closer semantically similar documents are grouped, the more distinct and useful is a map is.

Starting point is one interesting document, which the user has identified by unguided browsing or through keyword search depending on the value-added tools provided. Primarily, the user will evaluate documents close to this starting document on which he focused first. For that reason, the evaluation method uses one document as the starting point and calculates the Euclidean distance to all others. By sequentially using all documents as starting point, a similarity matrix is obtained for each mapping method. The ranked lists can be compared in order to determine the degree of correlation between the methods. This resembles the comparison of two similarity matrices row by row. The correlation between two ranked lists was measured using the Spearman coefficient [Hartu84]:

$$r = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n(n^2-1)} \quad [1]$$

r	correlation
n	number of items
d	distance in ranks between position of item in both lists

The similarity correlation measure was calculated for each row. The average of all rows was then calculated as the correlation between the matrices.

Thus, two maps could be compared. The results presented by [Mandl01] showed, that the correlation between maps created with SOM and Singular Value Decomposition were very low, thus, the maps are different.

This method does not question the general usability of these objects displays, however, it is helpful to answer several questions when constructing them:

- Do two methods for dimensionality reduction result in significantly different maps or environments, or are they very similar and only one method needs to be considered? In this case, the choice would be arbitrary.
- In the case of a comparative study, are two displays significantly different and are they therefore good candidates for a test?
- Are the features of environments robust over different domains (e.g. software, text or multimedia documents) and corpora (e.g. sets of documents)?

Those are questions to be answered before a user study is carried out. They address diverse aspects of map design. By answering these questions it can be made sure that the maps tested are quite different from each other. Thus, the amount of user testing can be significantly reduced by the preceding analytical experiments described in this article. Additionally to the above mentioned test we introduced hypothetical user behavior in order to cross-check and to refine our results.

3.2 INCLUDING USER STRATEGIES INTO THE EVALUATION

In order to simulate user behavior we introduced two possible browsing strategies into our tests: the global and the local browsing strategy. Both are highly plausible strategies for user behavior.

- **Global browsing strategy:** This strategy assumes that the user browses through the environment starting at a central object, which may be highlighted after a search process or which is simply in the center of the display and the user focuses on it first. Furthermore, this method assumes that the user follows the main metaphor of spatial proximity and first browses objects close to the central object. Closeness can be calculated by a distance measure like the Euclidean distance. Therefore, the user's attention ideally extends like a growing circle in a 2D environment and like a growing globe in a 3D environment. After assessing a document, the user is assumed to return to the starting document and search for the next document from the starting point. Thus, the user takes a global point of view and never loses sight of his entry object.

- **Local browsing strategy:** Another strategy is the local strategy, where the user first navigates to the closest object and from there again to the closest object. The user always looks at the document closest to the last document if the new document has not been evaluated yet. For this strategy, the distance focus is local and based on the current document whereas for the first strategy, the focus is global and remains on the initial starting object. The second strategy resembles a chain of assessed documents whereas the first strategy results in a growing circle of interest.

As figure 2 shows, both strategies usually lead to a different order of assessment of the documents, when the sequential list in the order of assessment is derived.

The situation presented in figure 2 where the same map leads to different sequential evaluation of the documents, seems to be a common consequence of the different user strategies.

3.3 EXPERIMENTS AND RESULTS

The data chosen for the experiment consists of 1000 documents from the database of social science documents which are intellectually indexed. These data are part of the international initiative for the evaluation of multilingual information retrieval systems CLEF[5] (Cross Language Evaluation Forum, [Peter01]). CLEF provides a test bed for multilingual retrieval systems including documents, queries and relevance assessment.

5 http://www.clef-campaign.org

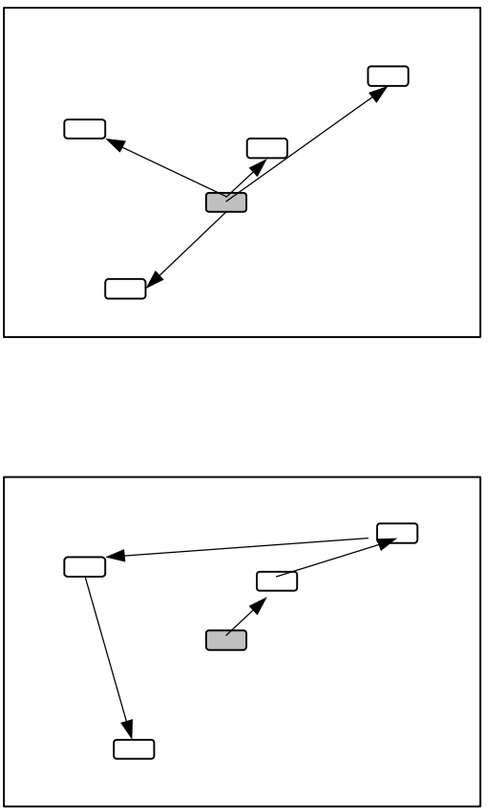

Figure 2: Global and local user strategy in identical maps

The documents are the objects which need to be visually organized. The terms assigned to them manually are the properties according to which their semantic distances were calculated. The property values are binary.

From this document set, two methods for the creation of two-dimensional displays were compared. One is based on latent semantic indexing (LSI, using software provided by [Berry93]) and the other one on the Kohonen Self-Organizing map (SOM).

Latent semantic indexing is a mathematical algorithm based on singular value decomposition. LSI has been developed for information retrieval and its developers claim that it captures the most salient features of a sparse document term matrix. It creates a compressed form of the matrix [Berry93]. In our study we compressed the term space to two dimensions to be able to create a visualization.

The SOM is basically an iterative clustering method, where each cluster is represented by a neuron which is associated with a weight vector of the same dimensionality as the input space. Whenever a pattern is assigned to a cluster neuron, the weight vector is changed according to the new input pattern. The most important feature of SOM

clustering is that neighboring clusters are similar to each other. To assure that, each neuron which lies close to the winning neuron to which the pattern was assigned is also modified. The weight vectors of the neighboring neurons are modestly changed according to the values of the input pattern and their distance from the cluster neuron. ([Bose96], [Oja99]).

The results of [Mandl01] suggest that the size of the maps does not influence the results as long as many different documents sets are tested. Thus, testing documents sets of a several million documents and testing documents sets of only a few documents does lead to similar results as long as the documents belong to one and the same domain.

In the test presented here, small maps with either ten or twenty documents were constructed with both LSI and SOM. These smaller maps were chosen since a user commonly does not evaluate many items in a display. In order to create the maps, a subset of ten and in a second test twenty documents from the corpus was extracted. From these documents either algorithm calculated a two-dimensional map based on the term vectors of the documents. Since the maps are difficult to compare, we chose the perspective of the user for comparison. The user is satisfied when he or she encounters appropriate documents when scanning the map. The map is thus perceived as a sequential list of items which may differ from user to user. However, the distance between items is an important influence factor for the likelihood for a document being evaluated.

In order to create the sequential ranked list of documents from the map, we assumed users relied either on a local or a global browsing strategy. The correlation between the maps with identical documents were calculated as the correlation between the ranked lists of the resulting lists. By this means, the influence of both the algorithms and the browsing strategy were compared.

	SOM local strategy	SOM global strategy	LSI local strategy	LSI global strategy
SOM local strategy	-			
SOM global strategy	**0.516**	-		
LSI local strategy	**0.076**	**0.01**	-	
LSI global strategy	**0.056**	**0.004**	**0.514**	-

Table 1: Correlation for 100 maps with 10 documents

To achieve more reliable results, 100 maps with ten as well as 100 maps containing twenty documents were created for each algorithm. The results are presented in table 1 and table 2.

The tendency in table 1 for the maps with ten documents and in table 2 for twenty documents is the same.

	SOM local strategy	SOM global strategy	LSI local strategy	LSI global strategy
SOM local strategy	-			
SOM global strategy	**0.406**	-		
LSI local strategy	**0.147**	**0.052**	-	
LSI global strategy	**0.029**	**0.066**	**0.435**	-

Table 2: Correlation for 100 maps with 20 documents

Overall, correlations of 0.4 or 0.5 show that the maps are related, but not very similar. Values below 0.2 indicate that the maps are very different.

Surprisingly, the influence of the browsing strategy is much lower than the influence of the creation algorithm. Whereas there is no similarity between maps created with SOM or LSI, there is a considerable degree of correlation between a map being browsed with local and global strategy. This is true for both algorithms and for both sizes of maps tested. For users who stick to one strategy, the difference is quite significant. Table 3 summarizes the results. The average shows a significant difference.

	100 x 10 Documents	100 x 20 Documents
average correlation between strategies	**0.515**	**0.420**
average correlation between algorithms	**0.040**	**0.106**

Table 3: Comparison between the effect of user strategies and algorithms

This is quite surprising as usually, the users' properties or the human factors are considered to have a very high influence on final satisfaction of the user. In the case of two-dimensional displays, the user strategy plays an important role. However, we have to pay very much attention to choose the correct mapping, which may play an even more important role.

3.4 FURTHER EXTENSIONS OF THE EVALUATION METHOD

So far, the method has been applied only to 2D environments. However, Kohonen Maps and other methods can be used to create 3D environments as well [Schat98]. The evaluation method can be extended and adapted for 3D environments. In this case, the global strategy would result in a globe. Interaction mechanisms may bias some directions in a 3D environment. Especially, it may be difficult to navigate in the third dimension and therefore, the assessment of objects close on the third axis may be less likely. This effect could easily be measured by recording user paths. In the case where the dimension reduction or the display algorithm do not take that into account, the evaluation method should model the attention area as an elliptic sphere with the starting object in the middle.

Although this method is user-oriented and provides a clear result, it has some shortcomings. Any 3D environment that is not displayed in an immersive virtual reality system like a head mounted display or a cave is actually displayed in 2D. This fact and the resulting limitations for perception, orientation and movement within the environment certainly have an impact on the question of usability. Furthermore, individual user paths may actually not be in accordance with the assumed distance-oriented browsing. However, by considering two possible interaction strategies and taking the average of the correlation measure for many starting points, this problem should have less impact on the validity of the results.

The evaluation method presented here considers typical user behavior in browsing interfaces in general and two-dimensional maps specifically. 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 behavior, it can be assumed, that users have a tendency to assess documents in the reading direction. Therefore, documents below and on the right side of the starting point are more likely to be considered consequently. A new evaluation method would need to take this into account. Only the n closest documents which are below or on the right side of the starting document are considered. In this case, *n* must be chosen rather low. Obviously, the orientation of the map has to be considered as well. Once a Kohonen SOM has been calculated, it can be rotated without changing the mapping from the input space into the Kohonen space.

4. OUTLOOK

The evaluation method presented here is concerned with browsing interfaces created by the distortion of similarity spaces into a smaller dimensionality. The experiment shows that different algorithms result in the creation of extremely different maps in which the user is likely to perceive the documents in a different sequential order. That means, that the choice of an adequate algorithm for the domain is crucial for the success of any application including 2D maps.

The ultimate goal of this research would be to map individual browsing styles to graphic computation algorithms. This is certainly a long term goal and requires a lot of empirical studies with real users to support the findings.

Several extensions of the evaluation method are discussed which lead to the evaluation of 3D environments and other aspects.

The question remains, whether the Spearman coefficient is appropriate for the evaluation of information retrieval results where the items on tops positions in the ranked list are of higher importance. The position of very relevant documents will be crucial for the users' perception of the quality of retrieval results. Therefore, more elaborated correlation coefficients need to be developed which assign higher weight to items highly ranked in the result lists.

REFERENCES

[Berry93] Berry, M., Do, T., O'Brien, G., Krishna, V., Varadhan, S. (1993): SVDPACKC (Version 1.0) User's Guide. Technical report. Computer Science Department. University of Tennessee in Knoxville, USA. Retrieved November 30, 2000, from the World Wide Web: http://www.netlib.org/svdpack/svdpackc.tgz

[Bose96] Bose, N.; Liang, P. (1996): Neural Network Fundamentals with Graphs, Algorithms, and Applications. New York et al.

[Chen96] Chen, H.; Schuffels, C.; Orwig, R. (1996): Internet Categorization and Search: A Self-Organizing Approach. In: Journal of Visual Communication and Image Representation. 7(1). pp. 88-101.

[Eibl00] 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)

[Eib01] Eibl, M.; Mandl, T.; Stempfhuber, M. (2001): Metaphors vs. Visual Formalisms in Visual Information Seeking. In: Panhellenic Conference on Human-Computer Interaction (PC-HCI 2001) Workshop on Integrating Metaphors, Multimodality and Multimedia. Rio Patras, Greece. Dec. 8, 2001.

[Hartu84] Hartung, J. (1984): Lehr- und Handbuch der angewandten Statistik. München, Wien.

[Knepp97] Knepper, M.; Killiam, R.; Fox, K.; Frieder, O. (1997): Information Retrieval and Visualization using SENTINEL. In: Proceedings of the 7th Text Retrieval Conference Trec-7, pp.393-397. http://trec.nist.org

[Koske01] Koskela, M.; Laaksonen, J.; Oja, E. (2001): Self-organizing image retrieval with MPEG-7 descriptors. In: Ojala, T. (ed.): Infotech Oulu International Workshop on Information Retrieval (IR 2001). Oulu, Finland. Sept. 19.-21. 2001. pp. 151-160.

[Mandl01] Mandl, T.; Eibl, M. (2001): Evaluating Visualizations: A Method for Comparing 2D Maps. In: Smith, M.; Salvendy, G.; Harris, D.; Koubek, R. (eds.): Usability Design and Interface Evaluation: Cognitive Engineering, Intelligent Agents and Virtual Reality. Proc HCI Intl 2001 (9th Intl Conference on Human-Computer Interaction), New Orleans, August 2001. vol. 1. pp. 1145-1149.

[March95] Marchionini, G. (1995): Information Seeking in Electronic Environments.

[Merkl94] Merkl, D.; Tjoa, AM.; Kappel, G. (1994): Learning the Semantic Similarity of Reusable Software Components. In: Frakes, W.(ed.): Proc. of the Third International Conference on Software Reuse: Advances in Software Reusability. Rio de Janeiro, Nov. 1.-4. 1994. Washington et al. pp. 33-41.

[Mutsc01] 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.

[Niel99] Nielson, J. (1999): Heuristic Evaluation. http://www.useit.com/papers/heuristic

[Odgen98] Odgen, W.; Davis, M.; Rice, S. (1998): Document thumbnail visualizations for rapid relevance judgments: When do they pay off? In: Proc. of the 7th Text Retrieval Conference Trec-7, pp.599-612. http://trec.nist.org

[Oja99] Oja, E.; Kaski, S. (1999): Kohonen Maps. Elsevier.

[Ojala01] Ojala, T.; Kauniskangas, H.; Keränen, H.; Matinmikko, E.; Aittola, M.; Hagelberg, K.; Rautiainen, M.; Häkkinen, M. (2001): CMRS : Architecture for content-based multimedia retrieval: In: Ojala, T. (ed.): Infotech Oulu International Workshop on Information Retrieval (IR 2001). Oulu, Finnland. Sept 19.-21. 2001. pp. 179-190.

[Peter01] Peters, C. (2001): Cross Language Information Retrieval and Evaluation. Springer [LNCS]

[Raube01] Rauber, A.; Frühwirth, M. (2001): Automatically Analyzing and Organizing Music Archives. In: Constantopoulos, P.; Solvberg, I. (eds.): Proc. 5th European Conference on Research and Advanced Technology for Digital Libraries (ECDL 2001) Darmstadt Sept. 4.-8. Berlin et al.: Springer. pp. 402-414

[Rouss99] Roussinov, D.; Tolle, K.; Ramsey, M.; McQuaid, M.; Chen, H. (1999): Visualizing Internet Search Results with Adaptive Self-Organizing Maps. In: Hearst, M.; Frederic, G.; Richard, T. (eds.): Proc of the 22nd Annual Intl ACM SIGIR Conf. on Research and Development in Information Retrieval (SIGIR '99). Berkeley, CA Aug. 15-19. 1999.

[Swan98] Swan, R.; Allan, J.; Byrd, D. (1998): Evaluating a Visual Retrieval Interface: AspInquery at TREC-6. In: CHI `98 Workshop on Innovation and Evaluation in Information Exploration Interfaces. Los Angeles, April 1998. Retrieved Nov.30, 2000, http://www.fxpal.com/CHI98IE/submissions/long/swan/index.htm

[Schat98] Schatz, B. (1998): High-Performance Distributed Digital Libraries: Building the Interspace on the Grid. In: 7th IEEE Int Symp High-Performance Distributed Computing. pp. 224-234.

[Taska01] Taskaya, T.; Cotreras, P.; Feng; T.; Murtagh, F. (2001): Interactive Visual User Interfaces to Databases. In: Smith, M.; Salvendy, G.; Harris, D.; Koubek, R. (eds.): Usability Design and Interface Evaluation: Cognitive Engineering, Intelligent Agents and Virtual Reality. Proc HCI Intl 2001 (9th Intl Conference on Human-Computer Interaction), New Orleans, August 2001. vol. 1. pp. 913-918.

[Ye00] Ye, H.; Lo, B. (2000): A Visualised Software Library: Nested Self-Organising Maps for Retrieving and Browsing Reusable Software Assets. In: Neural Computing & Applications vol. 9(4). pp. 266-279.