Börner, Katy. (2000). Searching for the perfect match: A comparison of free sorting results for images by human subjects and by Latent Semantic Analysis, Information Visualisation 2000, Symposium on Digital Libraries, London, England, 19 -21July, pp. 192-197.

Searching for the perfect match: A comparison of free sorting results for images by human subjects and by Latent Semantic Analysis techniques

Katy Börner
Indiana University
School of Library and Information Science
10th Street & Jordan Avenue, Main Library 019
Bloomington, IN. 47405 USA
E-mail: katy@indiana.edu

Abstract

Information visualization can improve the access and manipulation of digital data if it conforms to human expectations, and supports the particular interface, task, data, and user. Detailed usability studies are required to prove this assumption. This paper reports a usability study of the data analysis algorithm applied in the LVis - Digital Library Visualizer that spatially visualizes search results derived from user queries of digital library collections. In particular, it presents a comparison of free sorting results for image data done by 20 human subjects and the data analysis result derived via Latent Semantic Analysis from the textual image descriptions. The results of the study suggest that people, if asked to categorize with no constraints, will use features that are often derivable from textual metadata. Human clustering strategies and resulting demands for image browsing systems are discussed and future work is outlined.

1. Introduction

Emerging digital libraries provide users with access to massive amounts of diverse types of digital information, increasing the need for more usable and efficient tools to extract and manage information. Digital libraries today, store not only textual but also image, audio, or even video data increasing the demands for powerful retrieval methods for multimedia databases.

Current image retrieval engines, for example IBM's Query By Image Content system, retrieve images based on visual image content properties such as color percentages, color layout, and textures occurring in the images. Often, content based queries are combined with text and keyword predicates. Some systems, for example AT&T's Shoebox, provide an audio annotation capability whereby users can speak about images. These audio annotations are automatically transcribed to text and are used during image retrieval. The closer the textual image descriptions match the features used by human subjects to retrieve images the higher the quality of the resulting retrieval result.

Oftentimes, retrieval engines return endless lists of rank-ordered matching documents or images. The preorganization of search results into categories of documents with similar content based on research on human clustering strategies seems to be a promising way to manage this information flood.

There exists a considerable body of research that aims at processing, analyzing, and visualizing large amounts of diverse online data to support people explore and navigate intuitively in a semantically organized information space (White & McCain, 1998; Small, 1999; Chen, 1999; Börner, SPIE 2000). However, there are very few studies that show how far the mathematically derived visualizations resemble the organization of images by their human users.

In psychology, the acquisition of painting style (as a basis for sorting) was studied by different researchers. In Hartley and Homa's (1981) studies, Ss initially classified 18 impressionist paintings, according to artist, into categories of 3-9 examples, followed by a transfer test given immediately or after a 2-wk delay. A measure of conceptual structure, derived from the multidimensionally scaled space, was found to be significantly correlated with transfer performance.

Sebrechts et al. (1999) conducted a usability study of their system NIRVE -- a tool that supports the visualization of search results. They reported that Ss overall liked the organizational aspects, including clustering of documents and the relational arrangements of clusters.

Chen and Dumais (2000) implemented an interface that organizes web search results into a hierarchical category structure using Support Vector Machine (SVM) classifiers. Usability tests revealed its qualitative (subject confidence and preference) and quantitative (task completion time etc.) superiority over a conventional search interface that arranges webpages in a ranked list. Future research will examine the degree of match between human categories and categories derived via SVM and the best presentation of search results in their category contexts among others.

The LVis - Digital Library Visualizer visually organizes image retrieval results semantically based on features derived from the textual descriptions via Latent Semantic Analysis (LSA) as described in section 4.3. This paper

reports an usability study conducted to compare the results of a free sorting task for image data done by 20 Ss and the LSA result to spatially layout images for browsing[1]. The paper starts with a description of the image data set used, describes the free sorting task conducted, and how its result compares with the LSA output. Finally, we discuss the cluster strategies applied by human subjects and Ss preferences for the organization and presentations of image search results to facilitate browsing. The paper concludes with an outlook.

2. Image dataset

The LVis - Digital Library Visualizer (Börner et al, 2000) aims to provide a visual 2-D and 3-D interface to the Dido Image Bank. Dido, a digital library at the Indiana University Department of the History of Art, stores about 9,500 digitized images from the Fine Arts Slide Library collection of over 320,000 images. It permits convenient access and use of images for teaching and research purposes via a web interface at http://www.dlib.indiana.edu/collections/dido/.

Images in Dido are stored together with a thumbnail representation of the image and a textual description:

```
BOSCH, HIERONYMOUS,
Garden of Earthly Delights,
Center.Full View, after 1500,
o/p 220x195 cm, Madrid, Prado.2823,
Saskia.llff-106a.1/10/88, RB.PTG.NTH
```

For the purposes of this test, four image data sets were retrieved from the Dido image data bank using the keywords "Bosch", "African", and two Chinese data sets "CHINA.PTG.5DYN OR CHINA.PTG.6DYN" as well as "CHINA.PTG.TANG". The sets contained 12, 17, 31, and 32 images respectively see Table 1. The data sets were chosen because of their variety of full and detailed views (Bosch), different art forms (masks, sculptures, photographs, paintings in African), and painting styles in CHINA.PTG) that encourage very different ways of sorting.

3. Free sorting task

In order to compare human image sorting results and data analysis result derived via LSA, subject have been ask to freely sort image data sets of image thumbnail presentations (approximately 40 x 35 mm) that were displayed randomly via a Java Applet on a desktop 19" monitor. Using the mouse Ss could pick, drag, and place images. Whenever an image is picked then its title is displayed above the image. Data sets have been presented such that order and sequence effects of task presentations are avoided.

During the test, each S had to freely sort one smaller (Bosch or African) and one larger China data set in a way such that similar images are placed close to each other. Ss had to press the "Start" button to start sorting (writing out the start time) and a "Done" button (writing out the end time as well as the final x-y positions of all images). After each sorting task Ss got a printout of their sorting result and have been asked to circle and label the categories they came up with.

Figure 1 shows the "African" data set as sorted and labeled by one of the subjects. Image 6 is picked and its title *"Bamileke.Cameroon.Mask in the form of an Elephant"* is shown. Without the title this particular image is hard to identify as a mask.

After the two sorting tasks were completed, Ss had to describe the categorization criteria they used, their preferred image organization scheme, problems they had, etc.

Ss have been asked to achieve a non-overlapping layout of images that shows the image numbers placed to the left of each image. However, Ss were instructed to sort the images in any way that they feel is appropriate for the image set, to come up with any number of categories, and to use any kind of labels.

Figure 1: Sorted "African" Data Set

Twenty Indiana University, Bloomington graduate students with various backgrounds including Information Science (70%), Telecommunications (15%), Music Theory (5%), computer science (5%), and Library Science (5%) participated in the experiment.

Exactly 40% of the Ss were male, 60% female. Participants' ages were recorded using ranges: 50% of the Ss were between the ages of 18 and 29, 45% between 30 and 39, and 5% between 40 and 49. 85% reported that they

[1] Of course, the closer the textual image descriptions match the features used by human subjects to retrieve and categorize images the higher the quality of the resulting organization of the retrieval result.

spent more than 5 hours per week on the Web and 65% spent more than 15 hours per week on the computer. Figure 2 shows the self-ratings of Ss knowledge in art, computers, data bases, and digital libraries is shown on a scale of 1= naive, ..., 5=expert. 85% of the subjects are right-handed and 15% left-handed. There are 65% native English speaker, 15% Chinese, and 20% Korean. 75% of the Ss touch type.

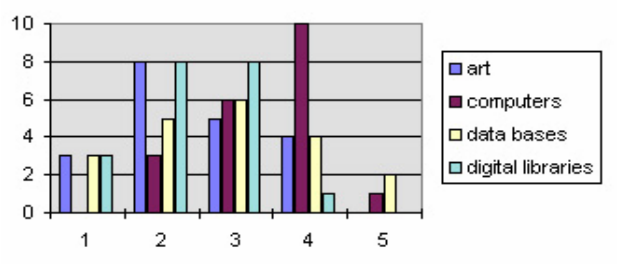

Figure 2: Knowledge Levels of Subjects

In the pre-test questionnaire Ss were asked if they prefer visual or textual information when they search the Web. 65% of the Ss wrote that they prefer visual information because it is 'easier to understand', 'aesthetically appealing', 'keeps me interested', 'more meaningful', 'gives direct information', 'entertaining', 'easy to identify', 'more intuitive', 'text based is quickly fatiguing'. 25% of the Ss stated that they use text and visual info and that they switch to text if images take too much time to load. Only 10% of the Ss stated that they prefer text because it is 'more direct'.

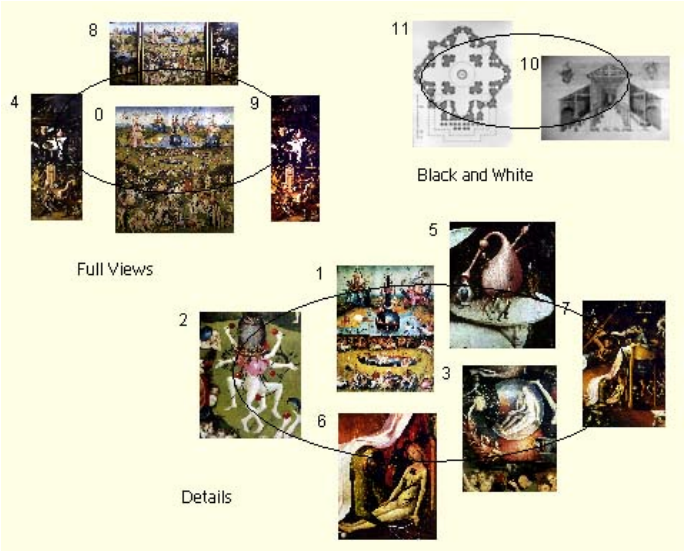

Figure 3: Sorted and Labeled "Bosch" Data Set

Figure 3 shows the sorted "Bosch" data set as sorted by one of the subjects. Again, the derived categories are labelled.

4. Results

The acquired data have been analyzed to derive information about the quality of the match between human free sorting results and LSA output as well as about the number and kind of categories formed or preferred by human subjects.

4.1. Number of categories and images per category

Based on the circled and labeled printouts of the sorting results, the average number of categories formed and the average number of images in each category were determined for all four data sets. The results are displayed in Table 1.

Data set	Bosch	African	TANG	DYN
Number of images in data set	12	17	31	32
Average number of categories	4.1	4.3	7	6.75
Number of images in category	2.8	3.8	4.29	5.42

Table 1: Data Sets Used and Results

A larger number of images seems to cause the formation of more clusters that contain more images. However, the small number of data sets used does not allow generalizations.

There have been noticeable differences between the time spent sorting by different subjects. It took some Ss almost twice as long to sort a particular image set. Follow-up experiments show that the Ss knowledge level of art influences the Ss sorting behavior as well as task completion times (see section Future Work).

4.2. MDS over subject data

A Multidimensional Scaling Analysis was performed to detect underlying dimensions in the Ss sorting data.

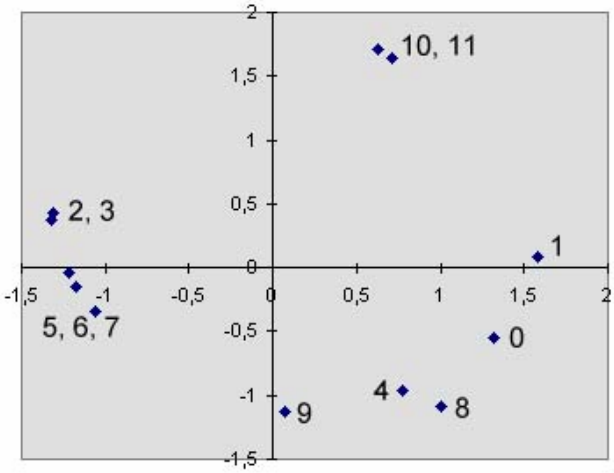

Figure 4: MDS Result for "Bosch"

Based on Ss grouping of images into categories, an image by image similarity matrix was derived in which each cell indicates how often the images (row) has been grouped in the same category with another image (column).

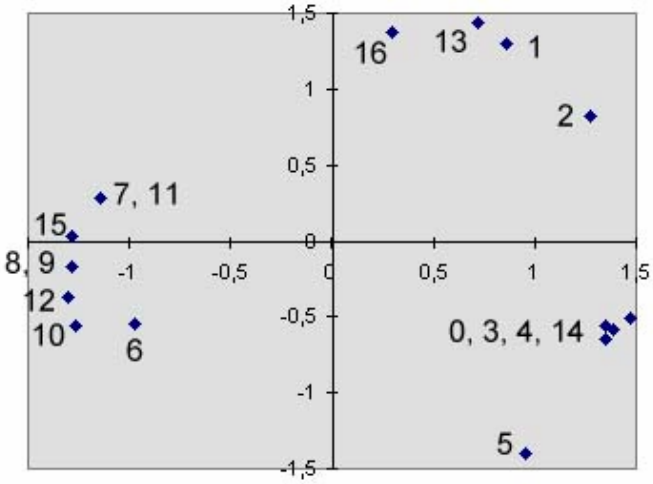

Figure 5: MDS Result for "African"

Figure 4 and 5 show the resulting MDS results for the "African" and "Bosch" data set. Each dot represents an image that is labeled by its number. Please have a look at Figure 1 and 3 to see the corresponding thumbnail images.

4.3. LSA over textual image descriptions

Latent Semantic Analysis was performed over the textual descriptions of the images in the four data sets. LSA is a corpus-statistic method for inducing and representing aspects of the meaning of words and passages (Berry, Dumais & O'Brien, 1995; Landauer, Foltz & Laham, 1998). Starting point of the analysis is a term-by-document matrix in which each cell indicates the frequency in which each term (rows) occurs in each document (columns).

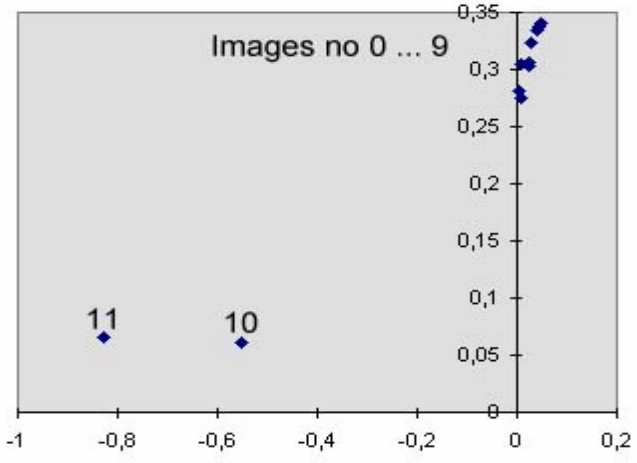

Figure 6: LSA Result for "Bosch"

LSA extends the vector space model by modeling term-document relationships using a reduced approximation for the column and row space computed by the singular value decomposition of the term by document matrix.

For the purposes of the comparison we derived a textual-image-description-by-image matrix for each of the four data sets.

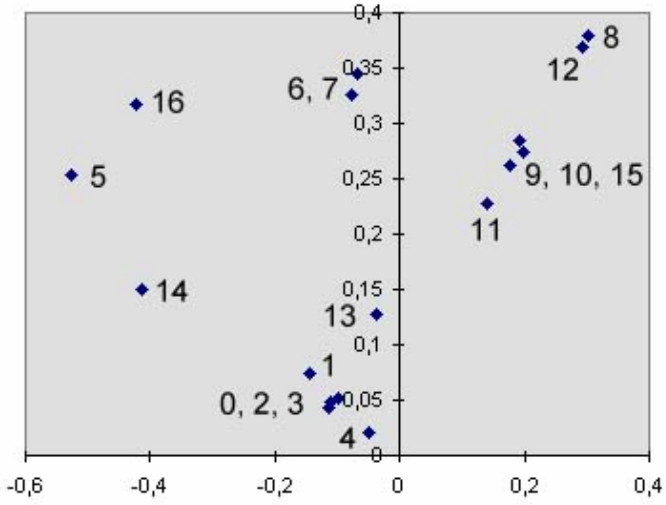

Figure 7: LSA Result for "African"

The SVDPACK by Michael Berry was used for the singular value decomposition. The rank-2 approximations for the row space of the textual-image-description by image matrix have been plotted for the "African" and "Bosch" data set in Figure 6 and 7.

4.4. Comparison

Although the plots for the MDS and LSA result look very different, a closer look reveals many similarities.

Comparing Figure 4 and 6, one can see that MDS and LSA results place images no. 10 and 11 close to each other and in a larger distance to the rest of the images showing Hironymous Bosch's *"Garden of Earthly Delights"* in full view (images 0, 1, 4, 8, 9) or detailed view (images 2, 3, 5, 6, 7).

As for the African data set shown in Figure 5 and 7, masks (images 6, 7, 8, 9, 10, 11, 12, and 15) are placed close to one another. Images 0, 3, 4, and 13 show outside scenery and are placed in spatial proximity in both diagrams. Remaining images such as the *"Ornament Bracelet"* (image 1), a *"Throwing Knife"* (image 13), or the *"Horse and Rider"* statue (image 16) have been grouped by Ss with very different other images and do not share many text commonalities.

The two Chinese data sets show a similar resemblance for subject and LSA data.

Images number 10 and 11 which show *"Michelangelo's Plan for completion of St. Peter's Vatican Plans (After Ferraboschi)",* cause a so called *extension effect* (Tversky, 1977). That is, images 0 ... 9 of *Bosch's Garden of Earthly Delights* become even more similar because of their dissimilarity to images 10 and 11. Their commonly shared

features such as image color and image content[2] are not shared by 10 and 11. These features therefore acquire *"diagnostic value"* and increase the similarity of images 0 .. 9 and thus reduce the distance between the corresponding dots in the plot.

In general, the MDS results reflect that Ss did spread images out in space – such that they don't overlap and all images are visible. LSA places images according to their similarity and pays no attention to the size of images.

The analysis of the Post-Test-Questionnaire provided several qualitative insights into Ss sorting strategies and preferences.

The answers for the question *"Which general categorization criteria did you use?"* as well as their frequencies (left hand side) are given below.

frequency	**used organization scheme**
12	what images depict, content, subject matter, depicted objects, topic image pattern, activities
6	(background) color
3	by their name, author, text label
2	what things are made of, materials
1	purpose objects were made for
1	image clarity
1	image size

Subjects answered the question *"How would you like images to be organized to facilitate your browsing?"* with

frequency	**preferred organization scheme**
13	image content
6	place of creation
9	creator
4	time
1	color
1	style
1	appearance

One subject suggested to organize images by four rated criteria: time (e.g. early impressionist), place (e.g., French), content (e.g., landscape) and author (e.g., Monet).

Asked *"Have there been too many or too few images on the screen?"*, 14 Ss reported that there were too many images on the screen, 3 reported too few, and 3 found it ok.

The question *"Did you have any problems to manipulate the images? If so, which?"* was answered as follows:

frequency	**problem**
4	selection of particular image especially when piled up
3	image size too small
2	no or identical text
1	overlapping of images
1	hard to see text

This supports the design decision for our final browser interface – it will provide a way to enlarge images and to see their full text description. The final interface uses a modified Boltzmann algorithm (Alexander et al, 1995) to spatially lay out semantic similarity networks for interactive exploration.

5. Conclusions

The results of the study suggest that people, if asked to categorize with no constraints, will use features that are often derivable from textual image descriptions.

In addition to this, the comparison gives some preliminary but encouraging results for the usage of LSA for data analysis. Clearly, LSA can only be applied to image libraries with good textual descriptions or metadata. Probably, certain image types, such as modern art, would be poorly categorized due to tendencies of representations not to relate very closely to peoples conception of the image. Annotation systems such as AT&T's Shoebox may provide a way to effortlessly generate extensive, content-based image descriptions that improve the quality of retrieval and categorization.

Last but not least the results comprise practical guidelines for selecting an appropriate number of images and of categories for display as well as for the preference of image organization schemes.

6. Future Work

Currently conducted usability tests with art experts and larger data sets show that the knowledge level in art significantly influences sorting behavior. Expert Ss typically sorted by painter and in rare cases by image content. They never sorted by color. A detailed comparison is under preparation.

Another line of usability studies compares sorting results using a 2-D desktop interface and a functionally equivalent 3-D immersive CAVE environment.

7. Acknowledgements

I would like to thank the students for participating in this study. Robert Goldstone has been a valuable discussion partner in the design and analysis of this study. The SVDPACK by Michael Berry was used for computing the singular value decomposition of large sparse matrices. I would like to give special tanks to Eileen Fry from Indiana University for her insightful comments on this research as well as ongoing discussions concerning the Dido Database. Thanks go to the anonymous reviewers for valuable recommendations on how to improve this paper. The research is supported by a High Performance Network Applications grant of Indiana University. Collaborators in this project are Andrew Dillon and Margaret Dolinsky.

[2] These features are frequently used by non-expert users to sort images, see below.

8. References

Alexander, Garcia, and Alder (1995) Simulation of the Consistent Boltzmann Equation for Hard Spheres and Its Extension to Dense Gases, Lecture Notes in Physics, Springer Verlag.

Berry, M. et al. (1993) SVDPACKC (Version 1.0) User's Guide, Univ. of Tennessee Tech. Report CS-93-194, (Revised October 1996). See also http://www.netlib.org/svdpack/index.html.

Berry, M, Dumais, S.T. & O'Brien, G.W. (1995) Using linear algebra for intelligent information retrieval. SIAM: Review, 37(4), 573-595.

Börner, K. (2000) Visible Threads: A smart VR interface to digital libraries. Proceedings of IST/SPIE's 12th Annual International Symposium: Electronic Imaging 2000, Visual Data Exploration and Analysis, San Jose, CA., pp 228-237, http://ella.slis.indiana.edu/~katy/SPIE00/

Börner, K. (in press) Extracting and visualizing semantic structures in retrieval results for browsing. ACM Digital Libraries, San Antonio, Texas, June 2-7, 2000. http://ella.slis.indiana.edu/~katy/DL00/

Börner, K., Dolinsky, M. & Dillon, A. (in press) LVis – Digital Library Visualizer. Information Visualization (IV2000). London, UK, July 19-21, 2000. http://ella.slis.indiana.edu/~katy/IV-LVis/

Chen C. (1999) Visualizing semantic spaces and author co-citation networks in digital libraries, *Information Processing and Management* 35, 401-420.

Chen, H. And Dumais, S. (2000) Bringing order to the Web: Automatically categorizing search results. Proceedings of the CHI 2000 Conference, pp 145-152.

Hartley, J. & Homa, D. (1981) Abstraction of stylistic concepts. Journal of Experimental Psychology: Human Learning and-Memory, Vol 7(1): 33-46.

QBIC(TM) -- IBM's Query By Image Content http://wwwqbic.almaden.ibm.com/

Landauer, T. K., Foltz, P. W., & Laham, D. (1998) Introduction to Latent Semantic Analysis. Discourse Processes, 25, 259-284.

Shoebox by AT&T Laboratories Cambridge http://www.uk.research.att.com/dart/shoebox/

Sebrechts, M., Vasilakis, J., Miller, M., Cugini, J. and Laskowski, S. (1999) Visualization of Search Results: A Comparative Evaluation of Text, 2D, and 3D Interfaces, In the Proceedings of the 22nd International Conference on Research and Development in Information Retrieval, pp. 3-10, M. Hearst, F. Gey, and R. Tong (eds), Berkeley, California.

Small, H. (1999) Visualizing science by citation mapping, *Journal of the American Society for Information Science*. 50(9):799-813.

The Indiana University Department of the History of Art Dido Image Bank, http://www.dlib.indiana.edu/ collections/dido/

Tversky, A. (1977) Features of similarity. Psychological Review. Vol 84(4):327-352.

White, H. D. and McCain, K. W. (1998) Visualizing a Discipline: An Author Co-Citation Analysis of Information Science, 1972-1995, *Journal of the American Society for Information Science*. 49(4):327-355.