tons on the edit page to help people learn the markup.

Creating the right conditions for a Wiki involves:

- setting up an effective initial structure, so that people can see where their contribution might fit;
- monitoring new and changed content, so that inappropriate content is dealt with promptly. This might also require clear guidelines on appropriate content, and a statement about the identity of the intended audience;
- having a statement about copyright and content ownership. If people sign their contributions, then it might not be considered appropriate for someone else to edit them; a comment (i.e., discussion or thread mode) might be better;
- providing an explicit Page History link to make it obvious that content can be restored if necessary;
- having basic text formatting tips displayed when someone edits a page, to help them remember the markup; and
- having suggested page naming conventions and writing style guidelines. The basic idea here is to remove some of the uncertainty associated with a totally open environment, which might help people overcome their initial hesitation about contributing.

In the early days, it might be useful to "shoulder tap" selected community members for specified Wiki contributions, to start the habit. Some Wikis have lists of "wanted pages" to identify topics they'd like someone to write about.

In some cases having a "comment on the Wiki" section encourages people to describe their reaction to the idea of editable Web content, and it can overcome their initial fear of breaking something if they edit a Wiki page.

Conclusion

The case studies in this paper illustrate just a few situations where Wikis can be effective. Because of their flexibility and simplicity, they can be used in a wide range of contexts, and provide an environment in which Berners-Lee's early vision for the Web can be achieved. A totally open Wiki might suit a widely distributed organization like the LIANZA ITSIG, while a members-only Wiki would suit a work group collaborating on policy documents or procedure manuals (or simply wanting an easy electronic notice board). Features like Recent Changes and Page History make it easy for community members to keep up with changes. Wikis are also gaining in popularity as simple, lightweight Web site CMSs for groups and individuals. Wikis offer libraries and other organizations a tool that can be used when upgrading traditional Web sites or implementing new Web-based projects—their potential for enabling Web-based communication with staff and users is just beginning to be appreciated.

Libraries were early and enthusiastic adopters of the Web as a medium to enhance and expand access to information for the communities they serve. The Web is constantly changing, and in the next few years we expect organizations to move to more interactive e-services. New technologies offer new opportunities and evolving Web accessibility standards present new challenges for libraries. For libraries looking to take advantage of new technologies and build Web sites that comply with Web accessibility standards, Wikis offer a relatively easy path to build the next generation of library Web sites. At the same time, they raise the possibility of having more interaction with users.

Finally, Wikis illustrate a shift to an increasing ability to use a Web browser as a person's main application tool, and they foreshadow other browser-based capabilities, such as table or drawing editors, which would make it possible to create complex documents using nothing more than a standard browser.

Author's note: all URLs in the text were accessed in early July 2005.

References and notes

1. Tim Berners-Lee with Mark Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web* (New York: HarperBusiness, 2000), 33.

2. *Wikipedia*, s.v. "Social software," June 30, 2005. Accessed July 1, 2005, http://en.wikipedia.org/wiki/social_software.

3. Bo Leuf and Ward Cunningham, *The Wiki Way: Quick Collaboration on the Web* (Boston: Addison-Wesley, 2001), 277.

4. Brian Lamb, "Wide Open Spaces: Wikis Ready or Not," *Educause Review* 39, no. 5 (Sept./Oct. 2004): 36–48.

5. Jeremy Frumkin, "The Wiki and the Digital Library," *OCLC Systems & Services: International Digital Library Perspectives* 21, no. 1 (2005): 18–22.

Graphical Table of Contents for Library Collections: The Application of Universal Decimal Classification Codes to Subject Maps

Victor Herrero-Solana, Félix Moya-Anegón, Vicente Guerrero-Bote, and Felipe Zapico-Alonso

The representation of information content by graphical maps is an extended

ongoing research topic. The objective of this article consists in verifying whether it is possible to create map displays using Universal Decimal Classification (UDC) codes (using co-classification analysis) for the purpose of creating a graphical table of contents for a library collection. The application of UDC codes was introduced to subject maps development using the following graphic representation methods: (1) multidimensional scaling; (2) cluster analysis; and (3) neural networks (self-organizing maps). Finally, the authors conclude that the different kinds of maps have slightly different degrees of viability and types of application.

Advanced techniques for information retrieval (IR) currently make up one of the most active areas of research in the field of library and information science. New models representing document content are replacing the classic systems in which search terms supplied by the user were compared against the indexing terms existing in the inverted files of a database. The objective of this article consists in verifying whether it is possible to create map displays using Universal Decimal Classification (UDC) codes, a classification system based on Dewey Decimal Classification, for the purpose of creating visualizations of a library collection.

One related topic of study in recent years is bibliographic browsing, a useful complement to querying strategies. Since the 1980s, a number of authors have dealt with this topic. For example, Ellis establishes that browsing is based on three different kinds of tasks: identification, familiarization, and differentiation.[1] Cove distinguishes three different browsing types: searching browsing, general purpose browsing, and serendipity browsing; whereas Bates presents six different types.[2] Yet most interesting is Bawden's browsing classification, which addresses similarity matching, structure-driven displays, and global vision.[3] Global-vision browsing implies the use of graphic representations, referred to in this article as *map displays*, that allow the user to grasp a global idea of the nature and structure of information in a database.

Several authors worked on this line of research throughout the 1990s, developing different types of maps. One of the most active authors was Lin, who introduced the concept of a graphical table of contents (GTOC) that is functionally analogous to the table of contents in the printed environment.[4] Lin applies the self-organizing map (SOM) algorithm to his own personal bibliography, analyzed by title and abstract fields, and represents it in a two-dimensional map.[5] The SOM algorithm is a major method for unsupervised learning, based on a grid of artificial neurons whose weights are adapted to match input vectors in a training set. It was first described by the Finnish professor Teuvo Kohonen and is thus sometimes referred to as a Kohonen map.[6] The algorithm takes a set of input objects, each represented by a vector in the matrix, and maps them onto nodes of a two-dimensional grid. Later on, Lin included such maps in the creation of GTOC Web sites based on a Java application.

Vectorization, the transformation of any information element into numerical data, using words from the title and abstract fields for co-word analysis, generates too large of a matrix, but this technique can be applied to reduced document sets. In this context, it is important to find some element that allows a less complex or "lighter" vectorization. Online public access library catalogs (OPACs) have certain elements, such as the subject codes of UDC, that can be more easily vectorized than free text in order to create GTOCs of a library collection.

Victor Herrero-Solana (victorhs@ugr.es) is Associate Professor and **Félix Moya-Anegón** (felix@ugr.es) is Professor, Department of Library and Information Science, University of Granada; **Vicente Guerrero-Bote** (vicente@alcazaba.unex.es) is Associate Professor and **Felipe Zapico-Alonso** (fzapalo@alcazaba.unex.es) is Associate Professor, Department of Computer Science, University of Extremadura, Badajoz, Spain. All of the authors are also affiliated with the SCImago Research Group (www.scimago.es/).

Materials and methods

The OPAC selected for this study is that of the Public Library of Granada, which contains 32,700 records and 43,900 UDC codes, an average of 1.34 codes per record. These records were vectorized using the UDC codes, so as to group them into twenty-seven major subject categories, derived from the hierarchical structure of UDC. The Pearson correlation index was applied to this matrix of data (27 x 32,700) to measure the similarities among these twenty-seven major classes and to generate a new matrix (27 x 27), to which the visualization method will be applied. This correlation index approach is widely used for science mapping construction.[7]

Two basic approaches were adopted in creating the display maps: (1) statistical (based on multivariate analysis); and (2) connectionist (usually, but not exclusively, based on artificial neural networks or ANNs).

Within the techniques of multivariate statistical analysis, three basic methods deserve mention at this point: (1) cluster analysis, (2) principal component analysis (PCA), and (3) multidimensional scaling (MDS).[8] According to Kinnucan, Nelson, and Allen, "These methods are referred to as dimensionality-reduction methods because this function is to simplify what might at first appear to be a complex pattern of associations among many entities."[9]

In the following sections, we review and summarize the characteristics of the three methods:

1. **Cluster analysis.** This technique is used to create two-dimensional

displays (e.g. dendrograms) of clusters of different objects whose relationships are represented by matrix values. This type of automatic classification, also known as numerical taxonomy, currently comprises more than 150 different techniques that are grouped in families according to shared procedures. Information science as a discipline generally involves polythetic clustering hierarchies, producing trees that illustrate the hierarchy of relationships among elements on the basis of individual characteristics.[10]

2. **Principal Component Analysis (PCA).** The basic premise of PCA is that the linear relation between any two variables is best summarized by a regression line. In other words, the variable that represents the regression line as a point cloud contains essential information about both variables. The two variables are thus combined into a single factor. This mechanism can be used to reduce pairs of variables to single dimensions in order to simplify the graphic display of the elements included in the matrix.[11]
3. **Multidimensional Scaling (MDS).** This multivariate analysis technique is used to identify the dimensions that best explain similarities and differences between variables. Because the purpose of MDS is to generate a map of objects, this approach can be considered an alternative to PCA.[12]

Neural networks are analytic techniques modeled after the (proposed) processes of learning in cognitive systems and the neurological functions of the brain. Neural networks use a data "training set" to build rules capable of making predictions or classifications on data sets. Neural networks can learn to assign multidimensional outputs to multidimensional inputs, and they do so while maintaining a great capacity for generalization. For this reason the better choice is the SOM algorithm. Kohonen's interest in discovering how an organization of this type might arise led him to investigate the subject.[13] The product of that research was the network model, bearing his name, that is capable of performing a topological organization of the inputs presented to it.

This type of network has recently been extrapolated to domain analysis, textual data mining, the extraction of semantic relationships among words in their contexts, and to the generation of topological maps of sets of documents, which may include labeling the zones of influence of each word or term.[14]

MDS maps

In the MDS-based display map each major subject category is placed in a certain point, depending on its relationship to other subject categories (figure 1). Also, each category is represented with a circle whose area is proportional to the volume of documents that it contains. The largest circles are located in the periphery, following the principle of center/periphery established by White and Griffith.[15]

The categories are classified in two large clusters: (1) science and technology, and (2) social science and humanities. This classification corresponds to its clustering, as in a Ward dendogram.[16] There are only two categories that do not seem to be in the expected cluster (economics and law); however, we should bear in mind that economics is related with categories of the science and technology cluster (mathematics). On the other hand, MDS places both categories at the edge of the map; in this way the dividing line can integrate them into the social sciences area.

SOM maps

The map display based on SOM is quite different from the MDS-based

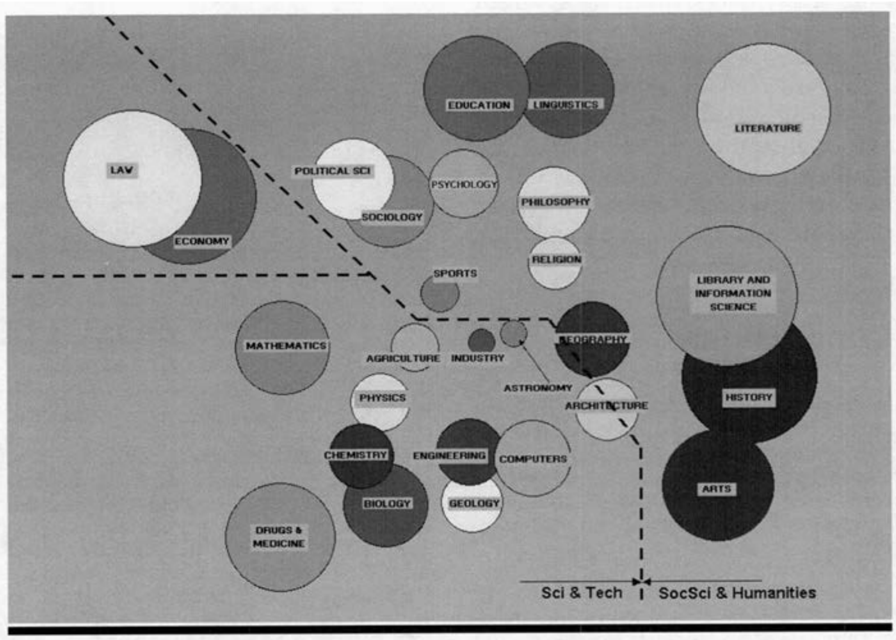

Figure 1. MDS map

map (figure 2). The SOM-based map is clearer, more schematic, and better ordered than the MDS-based map, but the size of each category is not proportional to the volume of documents, which might confuse the user. It is also very difficult to perceive the division of the categories in two big clusters as in the previous MDS-based map. The categories group according to neighboring relationships, a typical feature of Kohonen's algorithm, and the general structure of the collection that one could observe in the MDS-based map is much more difficult to discern here.

The neighboring relationships among the categories indicate the frequency of the co-occurrences of the classification codes. It is important to point out that the SOM searches out the best topology. This implies that when the representation must be reduced to two dimensions, the areas spread and the greater or lesser contact among them is indicative of the degree of interrelation. The proximity/distance among the areas is conditioned by co-classification frequency; however, this does not mean that the codes of classification of two categories that are far away from each other cannot co-occur at all.

To some extent the shapes of the areas are also determined by the co-classifications. These relationships cannot always be represented by the simplest geometric forms, for which reason their final appearance may strike the user as odd or unusual.

Conclusions

Despite the fact that user-based evaluation experience of this kind of map display is very limited, the following conclusions can be put forth:

- MDS and SOM are algorithms that can be used to generate bibliographic map displays
- An OPAC can be represented through co-classification analysis using UDC codes
- It is possible to use other decimal classifications, like DDC, but not Library of Congress
- MDS-based maps enhance viewing the structure of relations among the subject categories
- SOM-based maps are easy to use, because the view is clear, schematic, and well ordered
- SOM is easier to compute than MDS, especially when a lot of variables are involved

References and notes

1. D. Ellis, "A Behavioral Approach to Information Retrieval System Design,"*Journal of Documentation* 45, no. 3 (1989): 171–212.

2. J. F. Cove and B. C. Walsh, "Online Text Retrieval via Browsing," *Information Processing and Management* 24, no. 1 (1998): 31–37; M. Bates, "The Design of Browsing and Berrypicking Techniques for the Online Search Interface," *Online Review* 13, no. 5 (1989): 407–424.

3. D. Bawden, "Browsing: Theory and Practice," *Perspectives in Information Management* 3, no. 1 (1993): 71–85.

4. X. Lin, "Graphical Table of Contents," *Proceedings of the First ACM International Conference on Digital Libraries* (New York: ACM Pr., 1996): 45–53.

5. X. Lin, "Map Displays for Information Retrieval," *Journal of the American Society for Information Science* 48, no. 1 (1997): 40–54.

6. T. Kohonen, *Self-organizing Maps* (Berlin: Springer, 1997).

7. L. Egghe and R. Rousseau, *Introduction to Informetrics: Quantitative Methods in Library, Documentation, and Information Science* (Amsterdam: Elsevier, 1990); M. Kinnucan, M. Nelson, and B. Allen, "Statistical Methods in Information Science Research," *Annual Review of Information Science and Technology* 22 (1987): 147–78.

8. Egghe and Rousseau, *Introduction to Informetrics.*

9. Kinnucan, Nelson, and Allen, "Statistical Methods in Information Science Research."

10. H. D. Howard and K. W. McCain, "Visualizing a Discipline," *Journal of the American Society for Information Science (JASIS)* 49, no. 4 (1998): 327–55.

11. F. Moya-Anegón, E. Jiménez-Contreras, and M. D. L. Moneda-Corrochano, "Research Fronts in Library and Information Science in Spain (1985–1994)," *Scientometrics* 42, no. 2 (1998): 229–46.

12. Howard and McCain, "Visualizing a Discipline"; Moya-Anegón, Jimé-

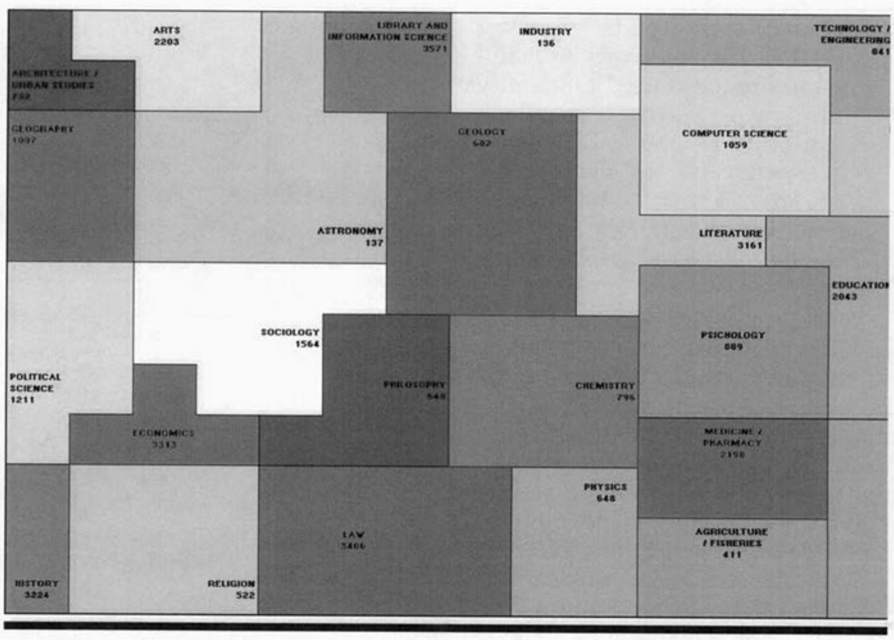

Figure 2. SOM map

nez-Contreras, and Moneda-Corrochano, "Research Fronts."

13. Kohonen, *Self-organizing Maps;* T. Kohonen, "Self-organized Formation of Topological Correct Feature Maps," *Biological Cybernetics* 43 (1982): 59–69.

14. H. D. White, X. Lin and K. W. McCain, "Two Modes of Automated Domain Analysis: Multidimensional Scaling vs. Kohonen Feature Mapping of Information Science Authors," *Structures and Relations in Knowledge Organization: Proccedings of the Fifth International ISKO Conference* (Wurzberg, Germany: Ergon Verlag, 1998); S. Kaski, K. Lagus, T. Honkela, and T. Kohonen, "Statistical Aspects of the WEBSOM System in Organizing Document Collections," *Computer Science and Statistics* (Fairfax Station, Va.: Interface Foundation of North America, 1998), 281–90; T. Honkela, V. Pulkki, and T. Kohonen, "Contextual Relations of Words in Grimm Tales, Analysed by Self-organizing Map," *Proceedings of the International Conference on Artificial Neural Networks ICANN-95* (Paris: EC2 et Cie, 1995); F. Moya-Anegón, V. Herrero-Solana, and V. Guerrero-Bote, "Virtual Reality Interface for Accessing Electronic Information," *Library and Information Research News* 22, no. 71 (1998): 34–39; F. Moya-Anegón et al., "NeuroISOC: un modelo de red neuronal para la representación del conocimiento," *Actas del IV Congreso ISKO-España EOCONSID '99* (Granada, Spain: ISKO-España, 1999), 151–56; V. Guerrero-Bote and F. Moya-Anegón, "Reduction of the Dimension of a Document Space Using the Fuzzified Output of a Kohonen Network,"*Journal of the American Society for Information Science and Technology* 52, no. 14 (2001): 1234–41.

15. H. D. White and B. C. Griffith, "Author Co-citation: A Literature Measure of Intellectual Structure," *Journal of the American Society for Information Science* 32, no. 3 (1981): 163–71.

16. Ward's method, based on the minimum variance principle, is a specific method used in cluster analysis for dendogram construction. A dendogram is a branching, tree-like diagram based on one or several criteria that can be used to illustrate the relationships between elements.