Chapter 4

Spatialization

André Skupin and Sara I. Fabrikant

Researchers engaged in geographic information science are generally concerned with conceptualizing, analyzing, modeling, and depicting geographic phenomena and processes in relation to geographic space. GI scientists consider spatial concepts, such as a phenomenon's absolute location on the Earth's surface, its distance to other phenomena, the scale at which it operates and therefore should be represented and studied, and the structure and shape of emerging spatial patterns. Geographic location is indeed a core concept and research focus of GI Science, and this is well reflected throughout the many chapters of this volume. In recent years, however, it has become apparent that the methods and approaches geographers have been using for hundreds of years to model and visualize geographic phenomena could be applied to the representation of any object, phenomenon, or process exhibiting spatial characteristics and spatial behavior in intangible or abstract worlds (Couclelis 1998). This applies, for example, to the Internet, in which text, images, and even voice messages exist in a framework called cyberspace. Other examples include medical records that have body space as a frame of reference, or molecular data structures that build up the human genome. These abstract information worlds are contained in massive databases, where billions of records need to be stored, managed, and analyzed. Core geographic concepts such as location, distance, pattern, or scale have gained importance as vehicles to understand and analyze the hard-to-grasp and volatile content of rapidly accumulating databases, from real-time stock market transactions to global telecommunication flows. This chapter is devoted to the use of spatial metaphors to represent data that may not be inherently spatial for knowledge discovery in massive, complex, and multi-dimensional databases. It discusses concepts and methods that are collectively referred to as *spatialization*.

What Is Spatialization?

In very general terms, spatialization can refer to the use of spatial metaphors to make sense of an abstract concept. Such spatialization is frequently used in everyday language (Lakoff and Johnson 1980). For example, the phrase "Life is a Journey" facilitates the understanding of an abstract concept ("human existence") by mapping from a non-spatial linguistic source domain ("life") to a tangible target domain ("journey") that one may have actually experienced in the real world. The desktop metaphor used in human–computer interfaces is another example for a spatial metaphor.

The role of spatial metaphors, including geographic metaphors, is also central to the more narrow definition of spatialization developed in the GI Science literature since the 19990s (Kuhn and Blumenthal 1996, Skupin and Buttenfield 1997, Skupin, Fabrikant, and Couclelis 2002), which is the basis for this chapter. Spatialization is here defined as the systematic transformation of high-dimensional data sets into lower-dimensional, spatial representations for facilitating data exploration and knowledge construction (after Skupin, Fabrikant, and Couclelis 2002).

The rising interest in spatialization is related to the increasing difficulty of organizing and using large, complex data repositories generated in all parts of society. Spatialization corresponds to a new, visual paradigm for constructing knowledge from such data. In the geographic domain, interest in spatialization stems largely from the growing availability of multi-dimensional attribute data originating from such sources as multi-temporal population counts, hyperspectral imagery, and sensor networks. New forms of data, still largely untapped by geographic analysis include vast collections of text, multimedia, and hypermedia documents, including billions of Web pages. A number of examples are discussed in this chapter highlighting the role of spatialization in this context.

The focus on spatial metaphors hints at a fundamental relationship between spatialization efforts and GI Science, with relevance beyond the geographic domain. Many spatio-temporal techniques developed and applied in GI Science are applicable in spatialization, and the ontological, especially cognitive, foundations underlying the conceptualization and representation of space can inform spatialization research. That is particularly true for a group of spatializations collectively referred to as "map-like" (Skupin 2002b), which are discussed and illustrated in some detail later in this chapter.

Spatializations are typically part of systems involving people exploring highly interactive data displays with sophisticated information technology. Most current spatialization research is directed at defining and refining various parameters of such interactive systems. However, the result of a spatialization procedure could also be a static hardcopy map that engages the viewer(s) in a discussion of depicted relationships, and triggers new insights (Skupin 2004). For example, one could visualize all the scientific papers written by GI scientists in 2006 in the form of a map printed on a large poster and use this to inspect the structure of the discipline at that moment in time. This can then encourage and inform the discourse on the state and future of the discipline much as a neighborhood map facilitates discussion on zoning ordinance changes during a city-planning forum.

Who Is Working On Spatialization?

The main challenge faced by anyone embarking on the creation of spatializations is that insights and techniques from numerous, and often disparate, disciplines need to be considered. *Visualization* research is very interdisciplinary and conducted by a heterogeneous group of loosely connected academic fields. *Scientific visualization* (McCormick, Defanti, and Brown 1987) and *information visualization* (Card, Mackinlay, and Shneiderman 1999) are two strands of particular interest for this discussion, both drawing heavily on computer science. The former is concerned with the representation of phenomena with physically extended dimensions (for example, width, length, height), usually in three dimensions. Typical application examples are found in such domains as geology (rock formations), climatology (hurricanes), and chemistry (molecular structures). Scientific visualization has obvious linkages with *geographic visualization* (see Chapters 11 and 16 of this volume, by Cartwright and Gahegan respectively, for two treatments of this topic) whenever the focus is on depicting phenomena and processes that are referenced to the Earth's surface. In contrast, information visualization is concerned with data that do not have inherent spatial dimensions. Examples include bibliometric data, video collections, monetary transaction flows, or the content and link structure of Web pages. Most information visualizations are in essence spatialization displays. Spatialization is thus best interpreted in the context of information visualization, which is quickly maturing into a distinct discipline, including dedicated conferences, scientific journals, textbooks, and academic degree programs.

Within GI Science, interest in spatialization tends to grow out of the *geographic visualization* community, which in turn mostly consists of classically trained cartographers. It is not surprising then that GIScientists involved in spatialization research draw inspiration from traditional cartographic principles and methods (Skupin 2000). On the other hand, ongoing developments in geographic visualization have also led to interactive, dynamic approaches that go beyond the static, 2D map (see Chapter 17 by Batty, in this volume, for some additional discussion and examples of this type) and within which spatialization tools can be integrated.

Data mining and knowledge discovery share many of the computational techniques employed in spatialization (see Chapter 19 by Miller, this volume, for some additional discussion of geographic data mining and knowledge discovery), for example artificial neural networks. Many preprocessing steps are similar, such as the transformation of source data into a multidimensional, quantitative form (Fabrikant 2001), even if these data sources are non-numeric.

Ultimately, spatialization is driven by the need to overcome the limited capacity of the human cognitive system to make sense of a highly complex, multidimensional world. That is why *psychology* and especially *cognitive science* have become influential disciplines in this research area. In this context it should be pointed out that while this chapter focuses on visual depictions, spatializations could include multimodal representations involving other senses such as sound, touch, smell, etc. In fact, the term spatialization first became known in the context of methods for producing 3D sound and detecting 3D spatial relationships from sound.

Computer science is still the dominant academic home to most spatialization efforts and has led the development of fundamental principles and novel techniques, especially in the human–computer interaction (HCI) field (Card, Mackinlay, and Shneiderman 1999). Few areas of scientific work have devoted as much effort to spatialization as information and library science, particularly when it comes to the analysis of text and hypermedia documents (Börner, Chen, and Boyack 2002, Chen 2003).

What Kinds Of Data Can Be Used For Spatialization?

Spatialization methodologies can be applied to many different types of data. One possible division of these would focus on the degree to which they are structured, leading to a distinction between structured, semi-structured, and unstructured data (Skupin and Fabrikant 2003). This is useful in terms of highlighting basic data transformation difficulties often encountered in spatialization. For example, unstructured text data may lack a clear indication of where one data item ends and another begins and can have dimensions numbering in the hundreds or thousands, as contrasted with multidimensional data typically used in geospatial analysis, where one rarely encounters more than a few dozen dimensions. However, given the focus of this volume on GI Science, this chapter considers two broad data categories. First, we discuss geographically referenced data, which are of obvious relevance to GI scientists. Then, much attention is given to data that are not referenced to geographic space or even related to geographic phenomena.

Geospatially referenced data

Why would one want to apply spatialization to *geographically referenced data* if cartographic depictions have proven useful for over 5,000 years and continue to be at the heart of current geovisualization research? Consider one very common example, the geographic visualization of demographic change. One almost always finds either juxtaposed maps of individual time slices or change condensed into composite variables (for example, relative percentage of growth). This may be sufficient for the visual detection of *change as such*, but does not easily support detection of *temporal patterns of change*. While location is what vision experts and cognitive psychologists call "pre-attentive" (MacEachren 1995, Ware 2000), this is basically taken out of play when geographically fixed objects, such as counties, are visualized in geographic space in this manner. Spatialization can eliminate that constraint by creating a new, low-dimensional representation from high-dimensional attributes. For example, one could take multi-temporal, multi-dimensional, demographic data for counties, map each county as a point and, with defined temporal intervals, link those points to form trajectories through attribute space (Skupin and Hagelman 2005). Thus, change becomes visualized more explicitly (Figure 4.1). One can then proceed to look for visual manifestations of common verbal descriptions of demographic change, such as "parallel" or "diverging" development (Figure 4.2). Traditional cartographic visualization in geographic space may also fail to reveal patterns and relationships that do not conform to basic assumptions about geographic space, such as those expressed by Tobler's First Law of Geography (Tobler 1970). With spatialization one can take geographic location out (or control for it) while focusing on patterns formed in n-dimensional attribute space.

In practice, spatializations derived from geographically referenced data will tend to be used not in isolation but in conjunction with more traditional geographic depictions. Due to their predominantly two-dimensional form, geometric data structures and formats used in GI Systems (GIS) are applicable to spatializations. They can be displayed and interacted with in commercial off-the-shelf GIS. Most examples

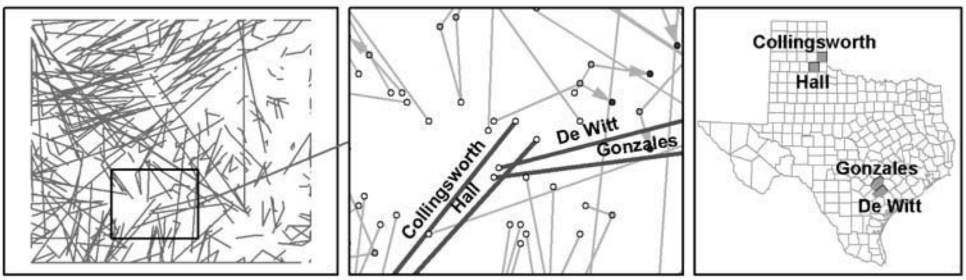

Fig. 4.1 Census-based visualization of trajectories of Texas counties based on data from 1980, 1990, and 2000 US population census
From Skupin and Hagelman 2005

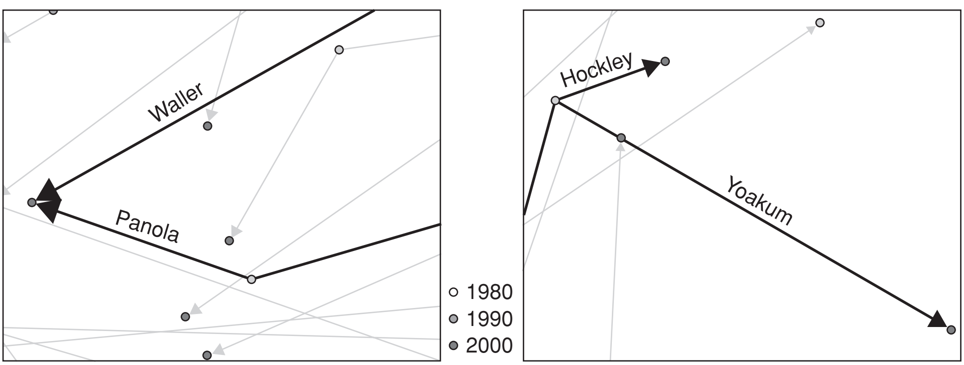

Fig. 4.2 Cases of convergence and divergence in a spatialization of Texas county trajectories
From Skupin and Hagelman 2005

shown in this chapter were in fact created in ArcGIS (Environmental Systems Research Institute, Redlands, California). Spatializations can also be juxtaposed to geographic maps, linked via common feature identifiers, and explored in tandem.

Many types of geographic data are suitable for spatialization. Population census data, for example, have traditionally been subjected to a number of multivariate statistics and visualization techniques, sometimes combined to support exploratory data analysis. Scatter plots and parallel coordinate plots (PCP) are established visual tools in the analytical arsenal. The spatialization methods discussed here do not replace these, but add an alternative view of multivariate data. In this context, it helps to consider how coordinate axes in visualizations are derived. In the case of the popular scatter plot method, each axis is unequivocally associated with an input variable. This is only feasible for a very limited number of variables, even when scatter plots are arranged into matrix form (Figure 4.3). Principal coordinate plots likewise exhibit clear association between axes and variables.

Contrast this with map-like spatializations, in which the relationship between input variables and display coordinates is far less obvious. Some even refer to the resulting axes as "meaningless" (Shneiderman, Feldman D, Rose A, and Grau 2000) and questions like "What do the axes mean?" are frequently encountered. They are difficult

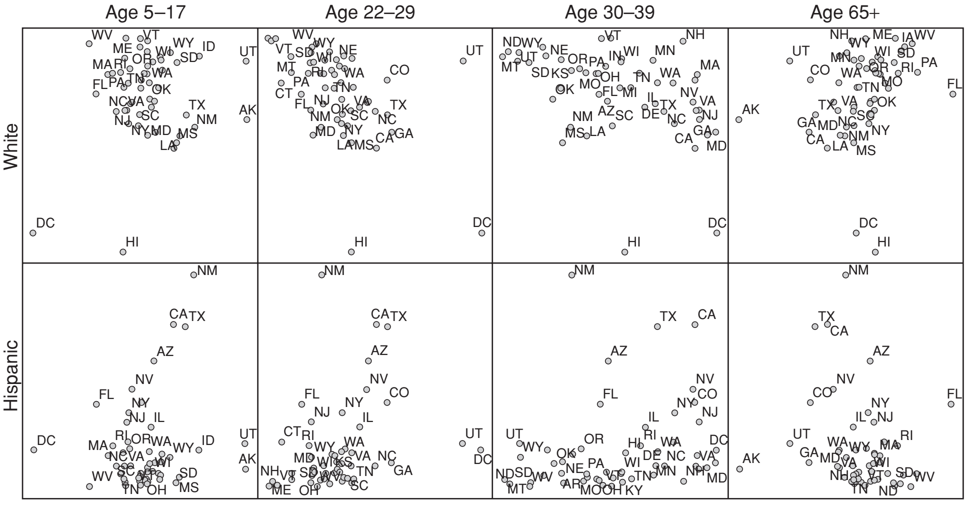

Fig. 4.3 Scatter plots derived from demographic data for US states

to answer, since in such techniques as multidimensional scaling or self-organizing maps *all* input variables become associated with *all* output axes. This allows a holistic view of relationships between observations (Figure 4.4). Figure 4.4 was derived by training an artificial neural network, specifically a self-organizing or Kohonen map (Kohonen 1995), with 32 input variables. Overall similarity of states becomes expressed visually through 2D point visualization. In addition, some of the input variables are shown as component planes in the trained Kohonen map to allow an investigation of relationships between variables.

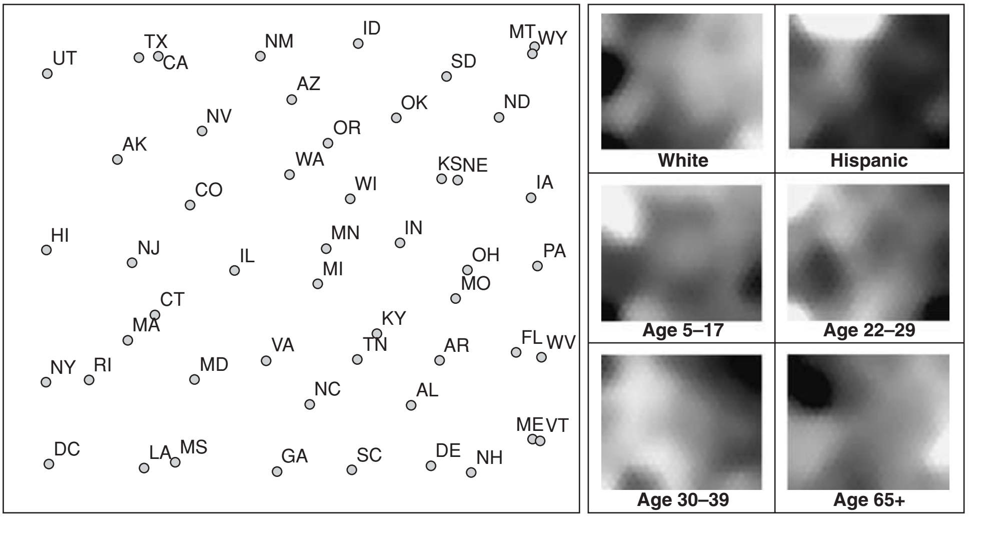

Fig. 4.4 Spatializations derived from 32 demographic variables using the self-organizing map method. Higher values in six (out of 32) component planes expressed as lighter shading

Data without geographic coordinate reference

Some of the most exciting and evocative developments in the visualization field in recent years have been efforts to apply spatial metaphors to non-geographic data or, more specifically, data that are not explicitly linked to physical space. Due to significant differences in how such data are stored, processed, and ultimately visualized, this section discusses a number of data types separately.

There are two broad categories of source data. One involves sources that already contain *explicit* links between data items, which in their entirety can be conceptualized as a graph structure. The goal of spatialization for this category is to convey such structures in an efficient manner in the display space. Hierarchical tree structures are especially common. A prime example is the directory structure of computer operating systems, like Windows or UNIX. Tree structures are also encountered in less expected places. For example, the Yahoo search engine organizes Web pages in a hierarchical tree of topics. The stock market can also be conceptualized as a tree, with market sectors and sub-sectors forming branch nodes and individual stocks as leaf nodes. Apart from such tree structures, data items could also be linked more freely to form a general network structure. The hypermedia structure of the World Wide Web is a good example, with Web pages as nodes and hot links between them. Scientific publications can also be conceptualized as forming a network structure, with individual publications as nodes and citations as explicit links between, generally pointing to the past. The exception might be preprints as they do not exist yet in their defining form. To illustrate this, we collected a few citation links from the *International Journal of Geographical Information Science* (*IJGIS*), starting with a 2003 paper by Stephan Winter and Silvia Nittel entitled “Formal information modelling for standardisation in the spatial domain.” The result is an origin–destination table of “who is citing whom” (Table 4.1). Later in this chapter, a visualization computed from this citation link structure is shown.

The second major group of non-georeferenced source data treats items as autonomous units that have no explicit connections among each other. Spatialization of such data relies on uncovering *implicit* relationships based on quantifiable notions of distance or similarity. This requires first a chunking or segmentation of individual data items into smaller units, followed by a computation of high-dimensional relationships. For example, the spatialization of text documents may involve breaking up each document into individual words. The following computations are then based on finding implicit connections between documents based on shared terms (Skupin and Buttenfield 1996). Similarly, images could be spatialized on the basis of image segmentation (Zhu, Ramsey, and Chen 2000). Other examples for spatializations involving disjoint items have included human subject test data derived from user tracking and elicitation experiments (Mark, Skupin, and Smith 2001).

How Does Spatialization Work?

The types of data to which spatialization can be applied are so heterogeneous that there really is no single method. As was stressed earlier, spatialization tends to draw on many different disciplines and integrating these influences can be challenging. For

Table 4.1 Origin–destination table for a citation network formed by papers within the *International Journal of Geographical Information Science*

Link	From Author	From Title	To Author	To Title
1	Su et al. (1997)	Algebraic models for the aggregation . . .	Abler (1987)	The National Science Foundation . . .
2	Su et al. (1997)	Algebraic models for the aggregation . . .	Rhind (1988)	A GIS research agenda
3	Su et al. (1997)	Algebraic models for the aggregation . . .	Brassel and Weibel (1988)	A review and conceptual framework . . .
4	Sester (2000)	Knowledge acquisition for the . . .	Su et al. (1997)	Algebraic models for the aggregation . . .
5	Lin (1998)	Many sorted algebraic data models . . .	Su et al. (1997)	Algebraic models for the aggregation . . .
6	Lin (1998)	Many sorted algebraic data models . . .	Kosters et al. (1997)	GIS-application development with . . .
7	Winter and Nittel (2003)	Formal information modeling . . .	Lin (1998)	Many sorted algebraic data models . . .
8	Lin (1998)	Many sorted algebraic data models . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
9	Clarke and Gaydos (1998)	Loose-coupling a cellular automaton . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
10	Shi and Pang (2000)	Development of Voronoi-based . . .	Okabe et al. (1994)	Nearest neighborhood operations . . .
11	Shi and Pang (2000)	Development of Voronoi-based . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
12	De Vasconcelos et al. (2002)	A working prototype of a dynamic . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
13	Cova and Goodchild (2002)	Extending geographical representation . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
14	Wu (2002)	Calibration of stochastic cellular . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
15	Wu and Webster (2000)	Simulating artificial cities in a GIS . . .	Takeyama and Couclelis (1997)	Map dynamics: integrating cellular . . .
16	Wu and Webster (2000)	Simulating artificial cities in a GIS . . .	Batty and Xie (1994)	Modelling inside GIS: Part I. model . . .
17	Wu and Webster (2000)	Simulating artificial cities in a GIS . . .	Peuquet and Duan (1995)	An event-based spatial temporal . . .
18	Wu and Webster (2000)	Simulating artificial cities in a GIS . . .	Burrough and Frank (1995)	Concepts and paradigms in spatial . . .
19	Wu and Webster (2000)	Simulating artificial cities in a GIS . . .	Clarke and Gaydos (1998)	Loose-coupling a cellular automaton . . .

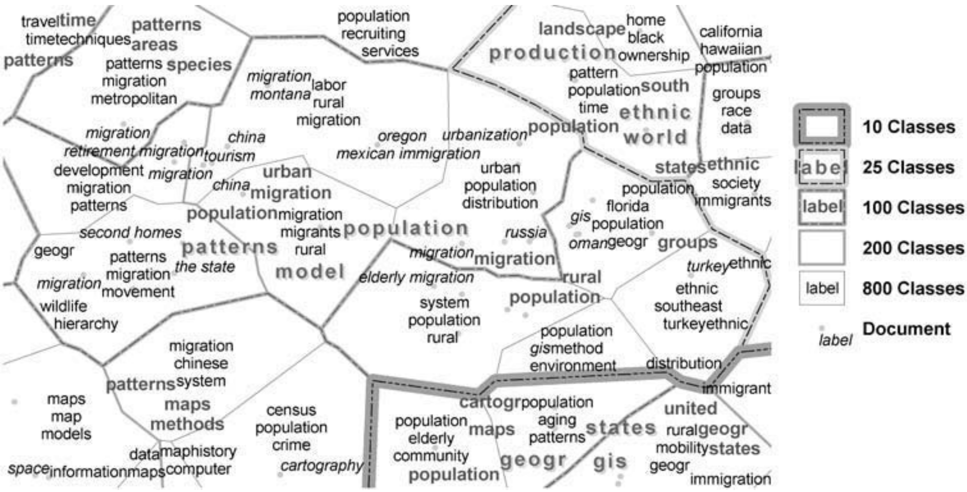

Fig. 4.5 Portion of a spatialization of conference abstracts. Five levels of a hierarchical clustering solution are shown simultaneously
From Skupin 2004

an example, consider the task of creating a map-like visualization of the thousands of abstracts that are presented at the annual meeting of the Association of American Geographers (AAG). This is an example of a *knowledge domain visualization* and would be useful in the exploration of major disciplinary structures and relationships in the geographic knowledge domain (Figure 4.5). Figure 4.6 shows the broad outline of a possible methodology for creating such a visualization. In the process, it also serves to illustrate the range of involved disciplines and influences, which include

- Information science and library science for creation of a term-document matrix, similar to most text retrieval systems and Web search engines (Widdows 2004);
- Computer science for the artificial neural network method used here (Kohonen 1995);
- GIS for storage and transformation of spatialized geometry and associated attributes;
- Cartography for scale dependence, symbolization and other design decisions.

Preprocessing

At the core of most spatialization procedures are techniques for dimensionality reduction and spatial layout. These tend to be highly computational, with very specific requirements for how data need to be structured and stored. Preprocessing of source data aims to provide this. In the case of well-structured, numerical data stored in standard database formats, preprocessing is fairly straightforward. For example, for single-year census data it will often involve only a few processing steps that can easily be accomplished using spreadsheet software, such as computation of z-scores, log transformations, or scaling of observations to fit into a 0–1 range.

The data to which spatialization is to be applied are, however, often not in a form that is amenable to immediate computation. In that case, much effort may

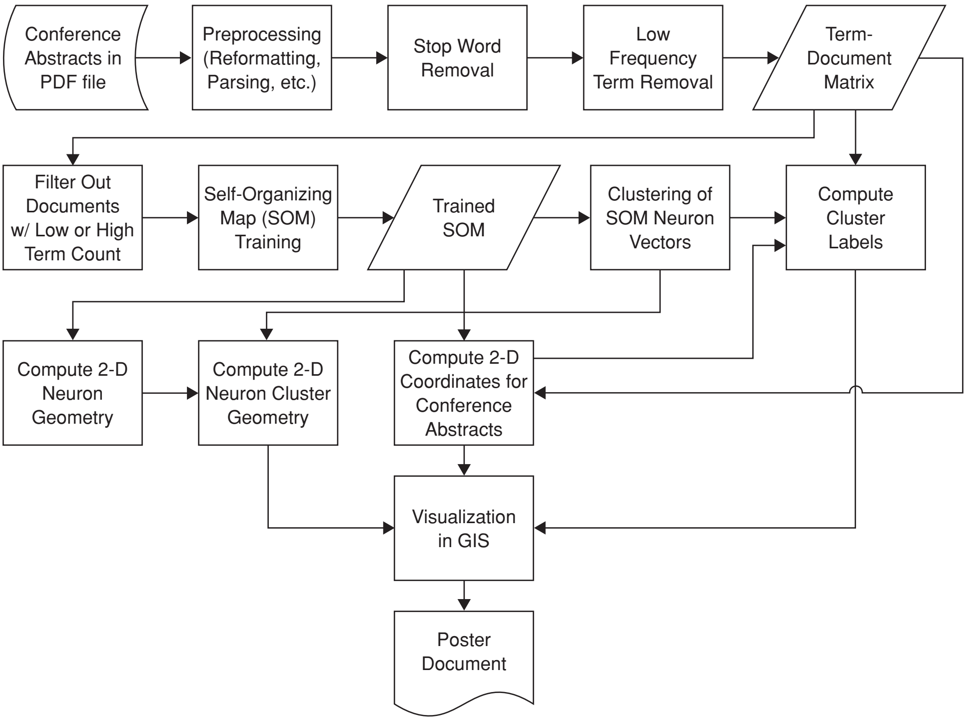

Fig. 4.6 Procedure for deriving a spatialization from AAG conference abstracts
From Skupin 2004

have to be devoted to reorganizing source data into a more suitable form. This can already be surprisingly difficult when dealing with multi-temporal, georeferenced data. Both geographic features and their attributes may be subject to change. For example, census block boundaries may be redrawn, ethnic categories redefined, and so forth. However, the resulting difficulties pale in comparison to source data in which there are no set definitions of what constitutes a feature, how features are separated from each other, or what the attributes should be that become associated with a feature.

What one is faced with here is a distinction between structured and unstructured data. The former is what one almost always encounters in GIS. Unstructured data present wholly different challenges. Consider the case of thousands of conference papers that one might have available in text form in a single file (Figure 4.7). There is no unequivocal separation between different documents nor clear distinction between content-bearing elements (title, abstract, keywords) and context elements (authors, affiliations, email addresses). One could look for certain elements (like end-of-line characters) useful for parsing, but such a procedure will be uniquely tailored to this particular data set, may suffer from inconsistencies in the data, and will require extensive modification to be used for differently organized data.

Semi-structured data are an attempt to address many of these problems by organizing data in accordance with a predefined schema. The extensible markup language

David Aagesen, Department of Geography, State University of New York, Geneseo, NY 14454. E-MAIL: aagesen@geneseo.edu. Still Dividing, Still Conquering: Conflict Over the Ralco Dam in Southern Chile.

The Bio-Bío River was the longest free-flowing river in southern Chile until the Pangue Dam was completed in 1997. Construction of a second dam, the Ralco Dam, is currently underway some ten kilometers upstream from the Pangue Dam. The reservoir to be created by the Ralco Dam, which proponents claim is necessary to ensure the efficiency and longevity of the Pangue Dam, will inundate nearly 3400 hectares and require the relocation of 85 indigenous Pehuenche families. The Ralco project has polarized the Pehuenche community in the upper Bío-Bio watershed. On the one hand, many Pehuenche consider Ralco a symbol of progress and embrace the project for its short-term employment opportunities. On the other hand, many Pehuenche view forced resettlement as a gross violation of their constitutional and human rights and have stated publicly that they will die fighting for their land if necessary. This paper outlines the evolution of territorial conflict in the upper Bio-Bío watershed in general, and conflict over the Ralco Dam in particular. It and weaken Pehuenche resistance. Material presented in this paper is based on fieldwork conducted in 1993 and a follow-up visit in July 2000. The paper includes a brief discussion of alternatives to the Ralco Dam that could satisfy energy demand in southern Chile without violating indigenous rights to land and resources.

Keyword: Chile, dams, indigenous geography.

Fig. 4.7 Conference abstract as unstructured text

(XML) is the most prominent solution to this. Figure 4.8 shows an example, in which a schema specifically designed for conference abstracts is applied to previously unstructured data. Such data offer many advantages. This XML file is suitable for human reading and computer parsing alike. From a software engineering point of view, this type of hierarchical, unequivocal structure is also very supportive of object-oriented programming and databases.

Spatialization depends on having data in a form that supports computation of item-to-item relationships in *n*-dimensional space. For structure-based methods, such as those based on citation links (see Table 4.1) or hypertext links, relationships are already explicitly contained and only have to be extracted to construct network graphs. For content-based analysis, the initial segmentation – for example the segmentation

```
<?xml version="1.0" encoding="iso-8859-1"?>
<CONFERENCE>
 <INDIVCONF>
  <CONFNAME>AAG 2001</CONFNAME>
  <YEAR>2001</YEAR>
  <CONFID>0012001</CONFID>
  <PLACE>New York</PLACE>
  <ABSTRACT>
   <ID>001200100001</ID>
   <TITLE>Still Dividing, Still Conquering: Conflict Over the Ralco Dam in Southern Chile</TITLE>
   <AUTHORINFO>
    <AUTHOR>
     <AUTHORID>00001</AUTHORID>
     <NAME>David Aagesen</NAME>
     <ADDRESS>Department of Geography, State University of New York, Geneseo, NY 14454</ADDRESS>
     <EMAIL>aagesen@geneseo.edu</EMAIL>
    </AUTHOR>
   </AUTHORINFO>
   <KEYWORDS>
    <KEYWORD>chile</KEYWORD>
    <KEYWORD>dams</KEYWORD>
    <KEYWORD>indigenous geography</KEYWORD>
   </KEYWORDS>
   <ABSTEXT> The Bío-Bío River was the longest free-flowing river in southern Chile until the Pangue Dam was completed in
        Construction of a second dam, the Ralco Dam, is currently underway some ten kilometers upstream from the Pangue
        Material presented in this paper is based on fieldwork conducted in 1993 and a follow-up visit in July 2000. The paper
   </ABSTEXT>
  </ABSTRACT>
```

Fig. 4.8 Conference abstract in semi-structured form as part of an XML file of a photograph or the identification of individual words within a text document – is followed by significant transformations (see top row in Figure 4.6). For example, text data may undergo stop word removal and stemming (Porter 1980, Salton 1989), as illustrated here:

INPUT: The paper includes a brief discussion of alternatives to the Ralco Dam that could satisfy energy demand in southern Chile without violating indigenous rights to land and resources
ONPUT: paper includ brief discuss altern ralco dam satisfi energi demand southern chile violat indigen right land resourc

From this, a high-dimensional vector can then be created for each document, with dimensions corresponding to specific word stems and values expressing the weight of a term within a document (Skupin and Buttenfield 1996, Salton 1989, Skupin 2002a).

Dimensionality reduction and spatial layout

The core of any spatialization methodology is the transformation of input data into a low-dimensional, representational space. In the case of data given as distinct features with a certain number of attributes one can rightfully refer to the corresponding techniques as *dimensionality reduction*. *Spatial layout* techniques are typically used when dealing with explicitly linked features, as in the case of citation networks.

Two popular dimensionality reduction techniques are *multidimensional scaling* (MDS) and the *self-organizing map* (SOM) method. MDS first requires the computation of a dissimilarity matrix from input features, based on a carefully chosen dissimilarity measure. Then, the method attempts to preserve high-dimensional dissimilarities as distances in a low-dimensional geometric configuration of features (Kruskal and Wish 1978). The popular Themescapes application (Wise, Thomas, Pennock, et al. 1995) is based on a variant of MDS (Wise 1999). Within GI Science, spatialization efforts have utilized MDS to create 2D point geometries for subdisciplines of geography (Goodchild and Janelle 1988), newspaper articles (Skupin and Buttenfield 1996, 1997), and online catalog entries (Fabrikant and Buttenfield 2001).

The SOM method is an artificial neural network technique (Kohonen 1995). It starts out with a low-dimensional (typically 2D) grid of *n*-dimensional neuron vectors. *N*-dimensional input data are repeatedly presented to these neurons. The best matching neuron to each observation is found and small adjustments are made to the vector of that neuron as well as to the vectors of neighboring neurons. Over time, this leads to a compressed/expanded representation in response to a sparse/dense distribution of input features. Consequently, major topological relationships in *n*-dimensional feature space become preserved in the two-dimensional neuron grid. One can then map *n*-dimensional observations onto it (left half of Figure 4.4), visualize individual neuron vector components (right half of Figure 4.4), or compute neuron clusters (Figure 4.5). SOMs have, for example, been used to spatialize Usenet discussion groups, Web pages (Chen, Schuffels, and Orwig 1996), image content (Zhu, Ramsey, and Chen 2000), conference abstracts (Skupin 2002a, 2004), and even

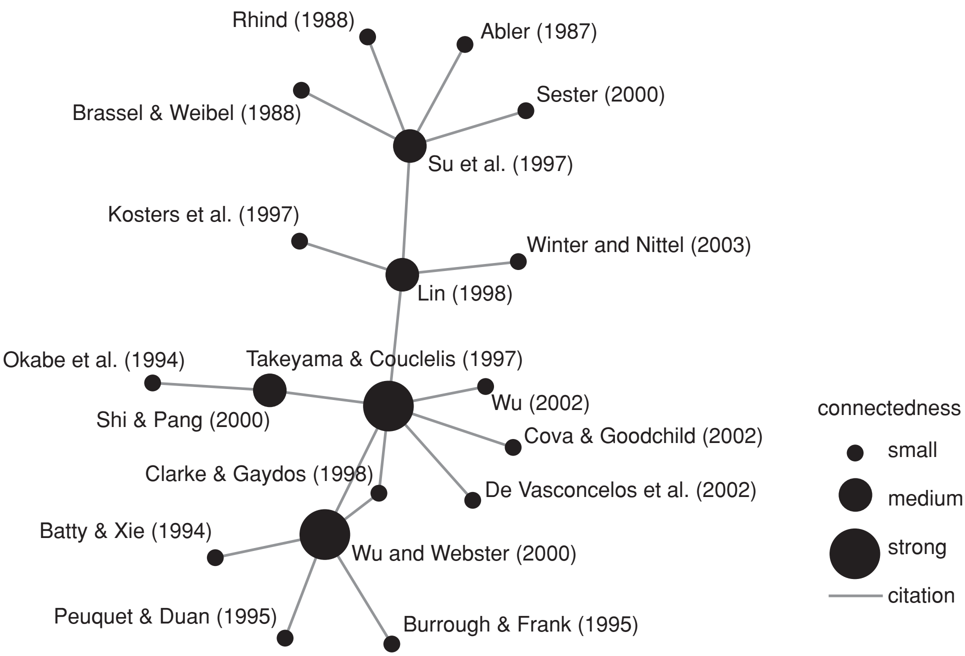

Fig. 4.9 Spring model layout and pathfinder network scaling applied to a small citation network formed by papers in the *International Journal of Geographical Information Science*

a collection of several million patent abstracts (Kohonen, Kaski, Lagus, et al. 1999). *Spring models* are another popular category of dimensionality reduction techniques (Kamada and Kawai 1989, Skupin and Fabrikant 2003).

Pathfinder network scaling (PFN) is a technique used for network visualization, with a preservation of the most salient links between input features. It is frequently applied to citation networks (Chen and Paul 2001). To illustrate this, we computed a PFN solution from the *IJGIS* citation data shown earlier. The result is a network structure consisting of links and nodes. When combined with a geometric layout of nodes derived from a spring model, the citation network can be visualized in GIS (Figure 4.9). Circle sizes represent the degree of centrality a paper has in this network, a measure commonly used in social network analysis (Wasserman and Faust 1999). Note how the centrality of the Takeyama/Couclelis paper derives from it being frequently cited (see Table 4.1), while the Wu/Webster paper establishes a central role because it cites a large number of *IJGIS* papers.

Among spatial layout techniques, the *treemap* method has become especially popular in recent years. It takes a hierarchical tree structure as input and lays portions of it out in a given two-dimensional display space (Johnson and Shneiderman 1991). In the process, node attributes can also be visually encoded (Figure 4.10). For example, when visualizing the directory structure of a hard drive, file size could be encoded as the area size of rectangles. Another important category are *graph layout* algorithms, which attempt to untangle networks of nodes and links in such a manner that crossing lines are avoided as much as possible and network topology is preserved.

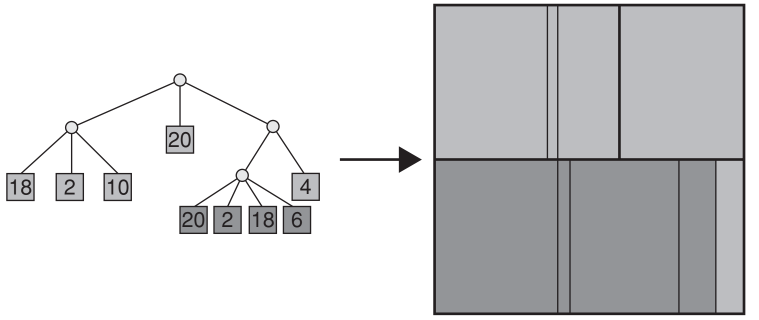

Fig. 4.10 The tree map method
From Skupin and Fabrikant 2003

Once dimensionality reduction or spatial layout methods have been applied, further transformations are necessary to execute the visual design of a spatialization. Depending on the character of the base geometry, these transformations may include the derivation of feature labels, clustering of features, landscape interpolation, and others (Skupin 2002b, Skupin and Fabrikant 2003). When dealing with 2D geometry, much of this can be accomplished in commercial off-the-shelf (COTS) GIS. Many aspects of these transformations remain to be investigated in future research, for instance how scale changes can be implemented as semantic zoom operations (Figure 4.11).

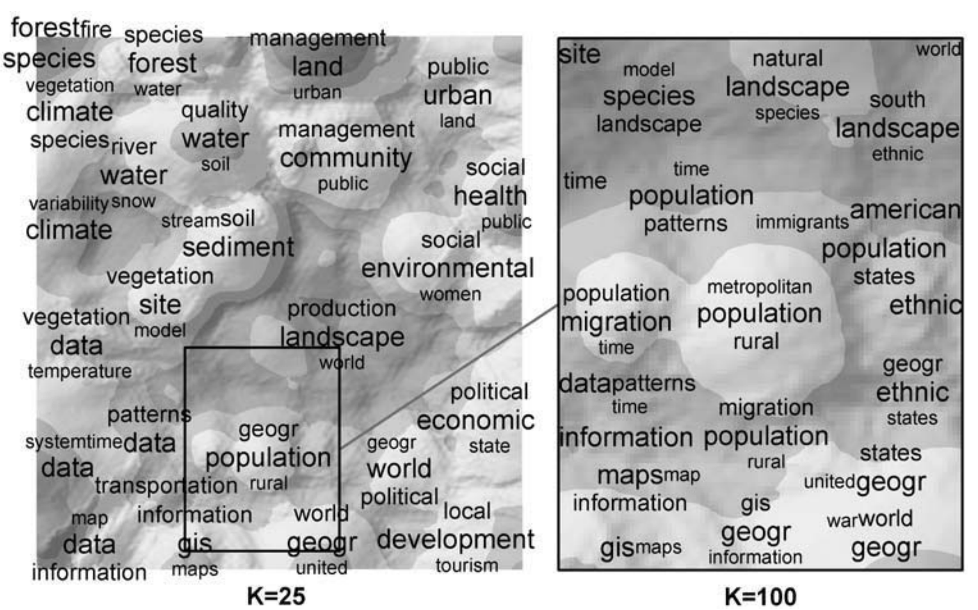

Fig. 4.11 Use of GIS in implementing scale-dependent spatialization of several thousand AAG conference abstracts. Labeling is based on two different *k*-means cluster solutions
From Skupin 2004

Spatialization geometry can also be linked to attributes that were not part of the input data set. For example, demographic change trajectories (Figures 4.1 and 4.2) could be linked – via symbolization or selection – to voting behavior or public policy decisions (Skupin and Hagelman 2005).

Usability and Cognitive Perspectives

An extensive set of display techniques has been developed for spatialization, and the impressive array of visual forms documents the productivity of this young academic field (Chen 1999). However, few researchers have succeeded in providing empirical evidence to support claims that interactive visual representation tools indeed amplify people's cognition (Ware 2000). Generally, non-expert viewers do not know how spatializations are created and are not told, through legends or traditional map marginalia, how to interpret such aspects of spatialized displays as distance, regionalization, and scale. Of the few existing experimental evaluations in information visualization, most evaluate specific depiction methods or types of software (Chen and Czerwinski 2000, Chen, Czerwinski, and Macredie 2000). While usability engineering approaches are good at testing users' successes in extracting information from a particular visualization, they do not directly assess the underlying theoretic assumptions encoded in the displays, the users' understanding of the semantic mapping between data and metaphor, and between metaphor and graphic variables, or the interaction of graphic variables with perceptual cues.

A fundamental principle in spatialization is the assumption that more similar entities represented in a display should be placed closer together because users will interpret closer entities as being more similar (Wise, Thomas, Pennock, et al. 1995, Card, Mackinlay, and Shneiderman 1999). Montello, Fabrikant, Ruocco, and Middleton (2003) have coined this principle the distance-similarity metaphor. For example, according to the distance-similarity metaphor, US states depicted in Figures 4.3 and 4.4 or conference abstracts shown in Figure 4.5 that are more similar to each other in content are placed closer to one another in the display, while spatialized items that are less similar in content are placed farther apart. In essence, this distance-similarity metaphor is the inverse of Tobler's (1970, p. 236) first law of geography, because similarity typically determines distance in spatializations. Thus we have referred to the "first law of cognitive geography" (Montello, Fabrikant, Ruocco, and Middleton 2003) – people believe that closer features are more similar than distant features. To the extent that this principle is true, it provides theoretical justification for the distance-similarity metaphor as a principle of spatialization design.

In a series of studies relating to point (Fabrikant 2001, Montello, Fabrikant, Ruocco, and Middleton 2003), network (Fabrikant, Montello, Ruocco, and Middleton 2004), region (Fabrikant, Montello, and Mark 2006), and surface display spatializations (Fabrikant 2003) Fabrikant and colleagues have investigated whether the fundamental assumption that spatialization can be intuitively understood as if they represent real-world spaces (Wise, Thomas, Pennock, et al. 1995, Card, Mackinlay, and Shneiderman 1999) is generally true. These studies provide the first empirical evidence of the cognitive adequacy of the distance-similarity metaphor in spatialization.

In these studies, participants have rated the similarity between documents depicted as points in spatialized displays. Four types of spatialization displays have been examined: (1) point displays (e.g., Figures 4.3 and 4.4), (2) network displays linking the points (e.g. Figures 4.1, 4.2 and 4.9), (3) black-and-white regions containing the points (e.g. Figure 4.5), and (4) colored regions containing the points (Figure 4.10). In the point displays, participants based judgments of the relative similarity of two pairs of document points primarily on direct (straight-line or "as the crow flies") metric distances between points, but concentrations of points in the display led to the emergence of visual features in the display, such as lines or clusters, that considerably moderated the operation of the first law of cognitive geography. In the network displays, participants based similarity judgments on metric distances along network links, even though they also had available direct distances across network links and topological separations (numbers of nodes or links connecting points). In the region displays, participants based similarity judgments primarily on region membership so that comparison documents within a region were judged as more similar than documents in different regions, even if the latter were closer in direct distance. Coloring the regions produced thematically-based judgments of similarity that could strengthen or weaken regional membership effects, depending on whether region hues matched or not. In addition, Fabrikant and Montello (2004) also gained explicit information on how similarity judgments directly compare to default distance and direct distance judgments. There are no differences between people's estimates of distance under default (nonspecified) and direct (straight-line) distance instructions for point, network, and region spatializations. Default distance instructions are interpreted as requests for estimates of direct distance in spatializations. They have also found that well-known optical effects such as the vertical (Gregory 1987) and space-filling interval illusion (Thorndyke 1981) affect distance judgments in spatializations and therefore may affect the operation of the first law of cognitive geography.

Without empirical evidence from fundamental cognitive evaluations the identification and establishment of solid theoretical foundations in spatialization will remain one of the major research challenges (Catarci 2000). A solid theoretical scaffold is not only necessary for grounding the information visualization field on sound science, but is also fundamental to deriving valid formalisms for cognitively adequate visualization designs, effective graphical user interface implementations, and their appropriate usability evaluation (Fabrikant and Skupin 2005).

Where Is Spatialization Going?

Spatialization addresses a need to make sense of the information contained in ever-growing digital data collections. There is considerable societal demand for the types of methods discussed in this paper. This includes such obvious applications as counter-terrorism work or the development of improved Web search engine interfaces. Telecommunications companies attempt to find patterns in millions of phone calls through spatialization. Private industry also hopes to use spatialization to detect emerging technological trends from research literature in order to gain a competitive advantage. Funding agencies would like to determine which research grant applications show the most promise. In recent years there have been a growing number of events dedicated to the type of research within which spatialization is prominently featured, organized by the National Academy of Sciences (Shiffrin and Börner 2004), the National Institutes of Health, the National Security Agency, and other public and private entities.

This chapter demonstrates that spatialization may be applicable to both georeferenced and non-georeferenced phenomena, whenever *n*-dimensional data need to be investigated in a holistic, visually engaging form. The involvement of GI scientists in spatialization activities does not have to be a one-way street in terms of using spatialization within particular applications. GI Science is also beginning to help answer fundamental questions with regards to how spatializations are constructed and used (Skupin, Fabrikant, and Couclelis 2002). Our understanding of cognitive underpinnings, usability, and usefulness is still quite incomplete. The computational techniques used for spatialization also need further investigation, especially when it comes to developing methods for integrated treatment of the tri-space formed by geographic, temporal, and attribute space. In summary, spatialization is an exciting area in which GI Science is challenged to address important issues of theory and practice for many different data and applications.

REFERENCES

Börner, K., Chen, C., and Boyack, K. W. 2002. Visualizing knowledge domains. In B. Cronin (ed.) *Annual Review of Information Science and Technology*. Medford, NJ: Information Today: 179–255.

Card, S. K., Mackinlay, J. D., and Shneiderman, B. 1999. *Readings in Information Visualization: Using Vision to Think*. San Francisco, CA: Morgan Kaufmann.

Catarci, T. 2000. What's new in visual query systems? In *Proceedings of the First International Conference on Geographic Information Science*. Savannah, GA, USA. Santa Barbara, CA: National Center for Geographic Information and Analysis.

Chen, C. 1999. *Information Visualisation and Virtual Environments*. London: Springer.

Chen, C. 2003. *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*. London: Springer.

Chen, C. and Czerwinski, M. 2000. Empirical evaluation of information visualizations: An introduction. *International Journal of Human-Computer Studies* 53: 631–5.

Chen, C., Czerwinski, M., and Macredie, R. D. 2000. Individual differences in virtual environments: Introduction and overview. *Journal of the American Society of Information Science* 51: 499–507.

Chen, C. and Paul, R. J. 2001. Visualizing a knowledge domain's intellectual structure. *IEEE Computer* 34(3): 65–71.

Chen, H., Schuffels, C., and Orwig, R. 1996. Internet categorization and search: A self-organizing approach. *Journal of Visual Communication and Image Representation* 7: 88–102.

Couclelis, H. 1998. Worlds of information: The geographic metaphor in the visualization of complex information. *Cartography and Geographic Information Systems* 25: 209–20.

Fabrikant, S. I. 2001. Evaluating the usability of the scale metaphor for querying semantic spaces. In R. D. Montello (ed.) *Spatial Information Theory: Foundations of Geographic Information Science*. Berlin: Springer Lecture Notes in Computer Science No. 2205: 156–72.

Fabrikant, S. I. 2003. Distanz als Raummetapher für die Informationsvisualierung (Distance as a spatial metaphor for the visualization of information). *Kartographische Nachrichten* 52: 276–82.

Fabrikant, S. I. and Buttenfield, B. P. 2001. Formalizing semantic spaces for information access. *Annals of the Association of American Geographers* 91: 263–80.

Fabrikant, S. I. and Montello, D. R. 2004. Similarity and distance in information spatializations. In *Proceedings of GIScience 2004*, Adelphi, MD, USA. Santa Barbara, CA National Center for Geographic Information and Analysis, pp, 279–81.

Fabrikant, S. I., Montello, D. R., and Mark, D. M. 2006. The distance-similarity metaphor in region-display spatializations. *IEEE Computer Graphics & Applications* 26: in press.

Fabrikant, S. I., Montello, D. R., Ruocco, M., and Middleton, R. S. 2004. The distance-similarity metaphor in network-display spatializations. *Cartography and Geographic Information Science* 31: 237–52.

Fabrikant, S. I. and Skupin, A. 2005. Cognitively plausible information visualization. In Dykes, J., MacEachren, A. M., and Kraak, M.-J. (eds) *Exploring Geovisualization*. Amsterdam: Elsevier: 667–90.

Goodchild, M. F. and Janelle, D. G. 1988. Specialization in the structure and organization of geography. *Annals of the Association of American Geographers* 78: 1–28.

Gregory, R. L. (ed.). 1987. *The Oxford Companion to the Mind*. Oxford: Oxford University Press.

Johnson, B. and Shneiderman, B. 1991. Treemaps: A space-filling approach to the visualization of hierarchical information structures. In *Proceedings of IEEE Visualization '91*, San Diego, CA, USA. Los Alamitos, CA: Institute of Electrical and Electronics Engineers, pp. 275–82.

Kamada, T. and Kawai, S. 1989. An algorithm for drawing general undirected graphs. *Information Processing Letters* 31: 7–15.

Kohonen, T. 1995. *Self-Organizing Maps*. Berlin: Springer-Verlag.

Kohonen, T., Kaski, S., Lagus, K., Salojärvi, J., Honkela, J., Paatero, V., and Saarela, A. 1999. Self organization of a massive text document collection. In E. Oja and S. Kaski (eds) *Kohonen Maps*. Amsterdam: Elsevier, pp. 171–82.

Kruskal, J. B. and Wish, M. 1978. *Multidimensional Scaling*. London: Sage University Paper Series on Qualitative Applications in the Social Sciences No. 07-011.

Kuhn, W. and Blumenthal, B. 1996. *Spatialization: Spatial Metaphors for User Interfaces*. Vienna, Austria: Technical University of Vienna.

Lakoff, G. and Johnson, M. 1980. *Metaphors We Live By*. Chicago, I: University of Chicago Press.

MacEachren, A. M. 1995. *How Maps Work*. New York: Guilford Press.

Mark, D. M., Skupin, A., and Smith, B. 2001. Features, objects, and other things: Ontological distinctions in the geographic domain. In D. R. Montello (ed.). *Spatial Information Theory: Foundations of Geographic Information Science*. Berlin: Springer-Verlag Lecture Notes in Computer Science No. 2205: 488–502.

McCormick, B. H., Defanti, T. A., and Brown, M. D. 1987. Visualization in scientific computing. *IEEE Computer Graphics and Applications* 7(10): 69–79.

Montello, D. R., Fabrikant, S. I., Ruocco, M., and Middleton, R. 2003. Spatialization: Testing the first law of cognitive geography on point-spatialization displays. In W. Kuhn, M. F. Worboys, and S. Timpf (eds) *Spatial Information Theory: Foundations of Geographic Information Science*. Berlin: Springer-Verlag Lecture Notes in Computer Science No. 2825: 335–51.

Porter, M. F. 1980. An algorithm for suffix stripping. *Program-Automated Library and Information Systems* 14: 130–7.

Salton, G. 1989. *Automated Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Reading, MA: Addison-Wesley.

Shiffrin, R. M. and Börner, K. 2004. Mapping knowledge domains. *Proceedings of the National Academy of Sciences* 101: 5183–5.

Shneiderman, B., Feldman, D., Rose, A., and Grau, X. F. 2000. Visualizing digital library search results with categorical and hierarchical axes. In *Proceedings of the ACM Conference on Digital Libraries (DL2000),* San Antonio, TX, USA. New York: Association of Computing Machinery, pp. 57–65.

Skupin, A. 2000. From metaphor to method: Cartographic perspectives on information visualization. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '00)*, Salt Lake City, UT, USA. Los Alamitos, CA: Institute of Electrical and Electronics Engineers, pp. 91–7.

Skupin, A. 2002a. A cartographic approach to visualizing conference abstracts. *IEEE Computer Graphics and Applications*, 22: 50–8.

Skupin, A. 2002b. On geometry and transformation in map-like information visualization. In K. Börner and C. Chen (eds) *Visual Interfaces to Digital Libraries*. Berlin: Springer-Verlag Lecture Notes in Computer Science No. 2539: 161–70.

Skupin, A. 2004. The world of geography: Visualizing a knowledge domain with cartographic means. *Proceedings of the National Academy of Sciences* 101: 5274–8.

Skupin, A. and Buttenfield, B. P. 1996. Spatial metaphors for visualizing very large data archives. In *Proceedings of GIS/LIS '96,* Denver, CO, USA. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, pp. 607–17.

Skupin, A. and Buttenfield, B. P. 1997. Spatial metaphors for visualizing information spaces. In *Proceedings of Auto Carto 13*. Bethesda, MD: American Congress of Surveying and Mapping/American Society for Photogrammetry and Remote Sensing, pp. 116–25.

Skupin, A. and Fabrikant, S. I. 2003. Spatialization methods: A cartographic research agenda for non-geographic information visualization. *Cartography and Geographic Information Science* 30: 99–119.

Skupin, A., Fabrikant, S. I., and Couclelis, H. 2002. Spatialization: Spatial metaphors and methods for handling non-spatial data. WWW document, http://www.geog.ucsb.edu/~sara/html/research/ucgis/spatialization_ucsb.pdf.

Skupin, A. and Hagelman, R. 2005. Visualizing demographic trajectories with self-organizing maps. *GeoInformatica* 9: 159–79.

Thorndyke, P. W. 1981. Distance estimation from cognitive maps. *Cognitive Psychology* 13: 526–50.

Tobler, W. 1970. A computer model simulating urban growth in the Detroit region. *Economic Geography* 46: 234–40.

Ware, C. 2000. *Information Visualization: Perception for Design.* San Francisco, CA: Morgan Kaufman.

Wasserman, S. and Faust, K. 1999. *Social Network Analysis*. Cambridge: Cambridge University Press.

Widdows, D. 2004. *Geometry and Meaning*. Stanford: CSLI Publications.

Wise, J. A. 1999. The ecological approach to text visualization. *Journal of the American Society for Information Science* 50: 1224–33.

Wise, J. A., Thomas, A. J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., and Crowet, V. 1995. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proceedings of the IEEE Symposium on Information Visualization (InfoVis '95),* Atlanta, GA, USA, pp. 51–8.

Zhu, B., Ramsey, M., and Chen, H. 2000. Creating a large-scale content-based airphoto image digital library. *IEEE Transactions on Image Processing* 9: 163–7.