This is a pre-publication draft only. For the final, published version, please refer to:
Skupin, A. (2007) Spatialization. in: Kemp, K. (Ed.) *Encyclopedia of Geographic Information Sciences*. Sage Publications.

Spatialization

Spatialization is the transformation of high-dimensional data into lower-dimensional, geometric representations on the basis of computational methods and spatial metaphors. Its aim is to enable people to discover patterns and relationships within complex *n*-dimensional data while leveraging existing perceptual and cognitive abilities. Spatialization can be applied to various types of data, from numerical attributes to text documents and imagery.

Spatial Metaphors

The main cognitive underpinning of spatialization lies in the extensive use of spatial metaphors, including cartographic and geographic metaphors, such as *map*, *scale*, *distance*, *region*, and so forth, which enable users to "see" *n*-dimensional relationships within a low-dimensional visualization. Empirical support for this approach is growing. For example, human subject studies have shown that people expect that distances between point symbols representing text documents correspond to the documents' relative similarity. Such distance-similarity relationships can further be manipulated through the introduction of network links (likened to cities connected by a road network) and through variations in symbology (e.g. the creation of visual groupings based on common symbols). In other words, the combination of the low-dimensional geometry with the set of traditional visual transformations developed largely within cartography can powerfully support the communication and discovery of complex, high-dimensional relationships.

Types of Spatialization Applications and Data

Spatialization can be applied to many different data types in many different application domains. Intelligence and law enforcement agencies want to sift through mountains of text, audio, and video content to extract useful nuggets of information. Research funding agencies receive thousands of grant applications every year, from which they would like to identify emerging research trends and promising new approaches. In an increasingly interdisciplinary research arena, both novices and experts find it difficult to stay abreast of the latest advances and are looking for condensed overviews of research results. Environmental scientists are increasingly receiving continuous streams of data from sensor networks, often involving dozens of attributes. Stock market analysts stand to gain from noticing unusual market movements, including relationships between different segments of the economy. Meanwhile, even average computer users find it difficult to keep track of the contents of hard drives containing hundreds of gigabytes of data.

Spatialization has been proposed as one possible approach for all of these scenarios. However, we are faced with extremely different data sources and the methods for turning them into a visual form vary accordingly. One way of distinguishing different data sources is according to the degree to which they already consist of a coherent, consistent structure that is supportive of further processing.

Some data are highly structured, with a clear distinction between different elements and consistent attribution, and stored in standard database tables. Population census attributes are a great example, as they are attached to distinct geographic features and have a consistent attribute structure (at least for a given census year). Environmental attributes collected by sensor networks are another example of well-structured data.

On the other hand, there are data exhibiting much less existing structure. Most text, audio, or image contents are good examples. The middle ground is occupied by semistructured data, in which a number of internal structural elements are already distinguished. XML (Extensible Markup Language) is the dominant form of semistructured data today, especially for text documents. Generally speaking, the more structured a data set is, the easier it will be to spatialize it.

Data sets to be spatialized can also be categorized according to how high-dimensional content and relationships are expressed in them. Many data sets are made up of individual elements with a set of *n* attributes corresponding to dimensions of an *n*-dimensional space. The notion of *dimensionality reduction* is readily applicable to such data. Examples are census enumeration units and associated attributes. Text documents can likewise be viewed as existing in an *n*-dimensional space, with dimensionality determined by the size of the vocabulary.

The second major type of data consists of elements that are explicitly linked to form network structures. Major characteristics of these structures derive from the specific application domain. For example, the files and folders on a hard drive form a hierarchical tree structure. Meanwhile, the networks formed by scientific articles are made up of unidirectional links, with citations always pointing towards past publications. Instead of indicating a specific location in high-dimensional space, network-type data mainly express topological relationships. These are meant to be preserved and conveyed in the low-dimensional display space and the corresponding methods are known as *spatial layouts*.

Methods of Dimensionality Reduction and Spatial Layout

Broadly interpreted, the term spatialization naturally implies turning something that is nonspatial into something that is spatial. To put it more accurately though, it addresses a transformation *between* different spaces whose main distinction is their dimensionality. People are used to physically navigating and manipulating three-dimensional space and are likewise able to make sense of visual, geographic representations of that space, which are typically derived from two-dimensional geometry – sometimes with added elevation values – and stored in GIS data bases. The goal of spatialization is to bring the same skills to bear on the complex, high-dimensional data sets that are produced throughout contemporary society. The result is a geometric representation in a low-dimensional, typically two- or three-dimensional space that is meant to enable people to detect patterns, relationships and trends inherent in high-dimensional data using perceptual and cognitive abilities that have proven so successful in dealing with geographic space.

Major examples of *dimensionality reduction* techniques commonly used for spatialization include multidimensional scaling (MDS), principal components analysis (PCA), spring models, and the self-organizing map (SOM) method. Among the factors to consider when choosing among dimensionality reduction methods are scalability (e.g., SOM is applicable to much larger data sets than MDS), performance (e.g., training of large SOMs can take a very long time), and distortion characteristics (e.g., MDS attempts to preserve certain *distance* relationships, while SOM results in the preservation of major *topological* structures).

Among *spatial layout* approaches, the tree map method has enjoyed particular popularity. It takes a tree structure as input and partitions a given rectangular area such that nodes are displayed as areas whose size and color express certain node attributes. Subdivisions are hierarchically stacked in accordance to the tree structure. This method has been popular for visualization of file/folder structures and of the stock market, where the tree structure derives from market sectors (e.g., banking stocks contained in the larger financial sector). There are also a number of graph layout methods, which arrange network nodes and links such that link intersections are largely avoided.

Spatialization Examples

Given the variety of data to which a spatialization strategy may be applied, two different examples are here presented, dealing with text contents and large network structures, respectively.

The first example is based on more than 20,000 abstracts submitted to the Annual Meetings of the Association of American Geographers (AAG) between 1993 and 2002. This example is quite illustrative of the long series of transformations that is frequently necessary in order to create a spatialization. In this case, it involved integration of heterogeneous, multi-temporal, text data, creation of a text index, filtering of the documents and vocabulary, creation of a self-organizing map, storage in a GIS database format, then a series of geometric and attribute manipulations, and finally visualization using GIS software. The result shows major topical subdivisions within this knowledge domain.

[Insert Spatialization_Figure_1 approx. here]

The second example is derived from papers published in the *International Journal of Geographical Information Science* (*IJGIS*). Let's assume that someone new to the field of GIScience – the very topic of this encyclopedia – would like to know who the most influential researchers in this field are and how they connect to form a scholarly research community. The most common approach would be to consult an expert in the field, another would involve going through a introductory text book on the subject and look for names mentioned there. Either way, the results will likely lack objectivity and comprehensiveness. Now, by far the most widely accepted expression of scholarly activity is the publication of original research articles in peer-reviewed journals, which we could presumably consult in our investigation of the GIScience community. The main impediment to using this resource is the sheer volume of published papers, which makes it impossible to actually survey the field manually. For example, a search of the *ISI Web of Science* database lists a total of 888 papers published between January 1991 and August 2006 in the leading GIScience journal, the *International Journal of Geographical Information Science*. Spatialization makes it possible to explore the GIScience knowledge domain based on an analysis of all of these papers, as seen in Spatialization_Figure_2. In this case, the more than 10,000 citations contained in the papers form a large network of citations (i.e., references from one paper to other papers) and co-citations (i.e., papers referenced by the same paper). Based on a combination of a spatial layout algorithm with measures of network connectivity one can derive visualizations of varying type and granularity. For example, one could derive networks of journals or networks of authors. The latter is analyzed here, with the large *IJGIS* citation network pruned to contain only the most influential authors and only the most dominant links among them. Node and font sizes express the total number of citations to a particular author.

[Insert Spatialization_Figure_2 approx. here]

Challenges in Spatialization

There is still a scarcity of spatialization tools that are at the same time easy-to-use, affordable for most users, and open to modification and integration. Loose-coupling of different programs is by far the most common strategy, with relative separation of data preprocessing, dimensionality reduction, geometric transformation, symbolization, and interaction.

Dealing with large non-structured, semi-structured, or inconsistently populated data set is a major impediment to creating useful spatializations. For example, note in Spatialization_Figure_2 how some authors appear more than once on the map, even when one is dealing with the same person, e.g., “Goodchild M” and “Goodchild MF”. This lack of author disambiguation is only one of the many challenges that are the subject of ongoing research by information scientists.

When dealing with geographically referenced data, spatialization of *n*-dimensional attributes should ideally be integrated with other methods of geographic visualization and spatial analysis. However, to date virtually none of the dimensionality reduction and spatial layout methods mentioned above have been integrated into commercial off-the-shelf (COTS) GIS software. Some programming toolkits now include spatialization modules that can be integrated into larger, exploratory visualization environments. The most common roles of COTS GIS in spatialization are the storage of two-dimensional geometry, execution of various geometric transformations (e.g., surface interpolation), and preparation of visual results.

As spatialization relies heavily on invoking spatial metaphors, much remains to be learned about the specific cognitive mechanisms involved and to derive from this parameters and guidelines for the effective design of spatializations.

Alternative Meanings of Spatialization

Note that spatialization is a term which outside of GIScience is sometimes used in different contexts. One is the use of sophisticated audio technology to create a spatial impression of the direction from which sound is emanating. This is also known as *sound spatialization*. Another notion of spatialization is encountered in the social sciences, where it refers to the process of uncovering the spatial character of human behavior, relations, and events (e.g., the gendering of space).

André Skupin

See also cartography; geovisualization; map metaphor; spatial cognition; spatial metaphor

Further Readings

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

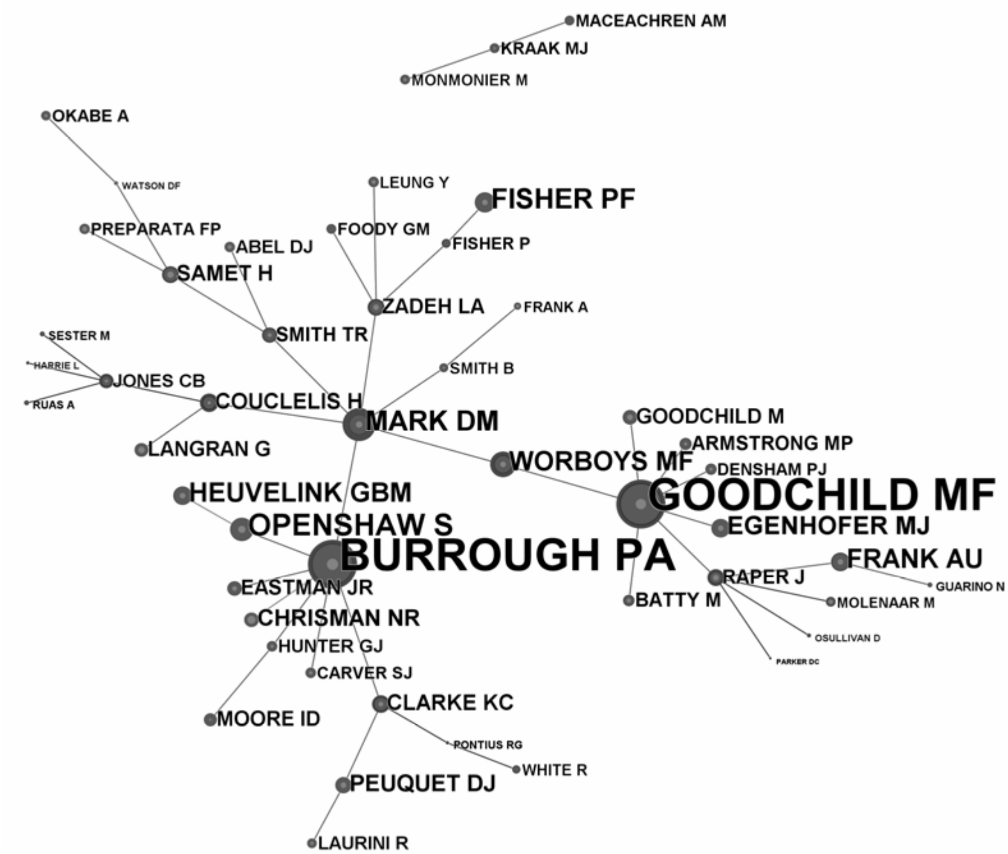
Couclelis, H. (1998) Worlds of Information: The Geographic Metaphor in the Visualization of Complex Information. *Cartography and Geographic Information Systems* 25: 209-220.

Fabrikant, S., and Skupin, A. (2005) Cognitively Plausible Information Visualization. in: Dykes, J., MacEachren, A., and Kraak, M.J. (Eds.), *Exploring Geovisualization*. Amsterdam: Elsevier. 667-690.

Skupin, A., and Fabrikant, S. (2003) Spatialization Methods: A Cartographic Research Agenda for Non-Geographic Information Visualization. *Cartography and Geographic Information Science*. 30 (2): 99-119.

Ware, C. (2004) *Information Visualization*. San Francisco: Morgan Kaufmann Publishers Inc.

Spatialization_Figure_1: Spatialization of the geographic knowledge domain based on 20,000+ abstracts submitted to the Annual Meetings of the Association of American Geographers (AAG).



Spatialization_Figure_2: Spatialization of a network of influential scientists derived from 10,000+ citations contained in the International Journal of Geographical Information Science (IJGIS).