Seminar Nasional
Aplikasi Teknologi Informasi 2004
Yogyakarta, 19 Juni 2004

The Role of Spatial Data Visualization in Knowledge Creation

Punang Amaripuja
Universitas Muhammadiyah Yogyakarta
Jl. Lingkar Barat, Tamantirto, Kasihan, Bantul, Yogyakarta 55183.
Telp. 0274-387656 ext. 118, HP. 0274-7471125
e-mail: p_amaripuja@yahoo.com

Abstrak

Proses pengubahan tacit knowledge menjadi explicit knowledge senantiasa menjadi batu sandungan bagi para ahli Knowledge Management. Kesulitan dalam 'memaksa' orang untuk men-transfer pengetahuan juga diikuti oleh kesulitan dalam memilih media yang paling tempat untuk melakukan transfer pengetahuan. Dalam makalah ini penulis berpendapat bahwa medium transfer yang paling tepat adalah data visual. Visualisasi dapat dikonstruksi dalam berbagai bentuk, semisal diagram, grafik dan peta. Data geografik diperkirakan terkandung dalam 85% dari semua database, sehingga data ini paling tepat di-transfer dalam bentuk visualisasi data spasial. Makalah ini selanjutnya menerangkan mengenai visualisasi data spasial dan peranannya dalam meningkatkan pemanfaatan intuisi dan kreativitas dalam knowledge creation. Beberapa contoh aplikasi.yang diberikan termasuk untuk data mining, geobrowsing, zoomable data interfaces, dan spatial data repositories.

Kata kunci: *knowledge management, knowledge creation, data spasial, visualisasi.*

1. Introduction

Knowledge Management has never been and will never be an exact science due to the abundance of variables involved. But Knowledge Management has never before been closer to its goal. The recent increase in computing power enables enterprises to simultaneously share and analyse huge amounts of data, which allows them to perceive data from a different perspective. At this stage, the limit of knowledge extraction and creation is mainly constrained by the ability of humans to comprehend information and knowledge in its various forms and sizes.

As a basis for our discussion, we will use the definition of knowledge coined by Sowa (1994, vide Cheah, 2002):"The knowledge of something is the ability to form a mental model that accurately represents the thing as well as the actions that can be performed on it and by it". This definition of knowledge emphasizes on the development of a mental model. The question remains on how to develop such a model that can be easily perceived and comprehended.

2. Knowledge Creation and Knowledge Discovery

Martin (1999, vide Cheah, 2002) suggests that the knowledge processes aspects of Knowledge Management are organizational structures, knowledge creation, resource-based-operation and performance measurement. He further describes knowledge creation as sharing, learning and transferring knowledge. Nonaka and Takeuchi (1995) offers an extensive study on knowledge creation, which they define as the capability to create new knowledge, disseminate knowledge throughout the organization, and embody knowledge in products, services & systems.

In order to support knowledge creation, the process of knowledge discovery has to be undergone. Fayyad et.al. (1996) explain that knowledge discovery could be aided by the methods he proposes: 1) *Classification*, the learning of a function that maps a data item into one of several predefined classes; 2) *Clustering*, the identification of a finite set of clusters to describe data; 3) *Visualization*, a description technique which is rich in picture concepts; and 4) *Regression*, the learning of a function that maps a data item into a real-valued prediction variable.

Knowledge creation and discovery will eventually support the decision-making process after the intelligent use of data, structured and unstructured knowledge, social processes and interactions among experts and non-experts. Leung (1997) summarizes this process in a model in which information consists of collection, representation, storage-retrieval, processing, and display of data. On the other side, knowledge consists of acquisition, representation and storage of knowledge, as well as the use of knowledge in inference and analysis.

3. Visualization as the Perfect Medium of Transfer

Knowledge sharing, discovery, and creation will only be achieved after the transfer of tacit knowledge into explicit knowledge. Tacit knowledge is deeply rooted in experience and is thus highly personal, subjective, and often intuitive. On the other hand, explicit knowledge is formal and systematic and usually expressed in words and numbers (Chan Cheah, 2002). These two types of knowledge are initially separated and without a suitable medium they will remain as such.

The ideal medium of transferring tacit knowledge into explicit knowledge would be a medium that can provide what Showa (1994) describes as a “mental model” that accurately represents tacit knowledge. Leung (1997) acknowledges the importance of representations and suggests that representation and storage are the two components that are present in both knowledge and information. This medium of transfer could either be classification, clustering, visualization or regression (Fayyad et.al., 1966).

In order to select the most ideal medium to reach our objective, we must note the characteristics of human reasoning. Leung (1997) suggests that humans often reasons with intuitions, values, experiences and judgement. He further explains that humans tend to organize knowledge with loosely defined concepts and structures. Thus, reasoning generally needs loosely structured common sense rather than highly structured mathematical models. In this sense, knowledge serves as a basis for inference and conversion of information into organized and understandable forms.

Visual representations perfectly match these criterions. Visual representations are unstructured and easily understandable, and they have the potential to ignite intuition and creative thinking. In doing so, visual representations can help meet one of the hardest challenges in Knowledge Management. However, they are not without drawbacks. Vekiri (2002) asserts that graphics are effective learning tools only when they allow readers to interpret and integrate information with minimum cognitive processing. Spatial data visualization, which will be further discussed, is a good solution to overcome this problem.

4. Spatial Data

Initially, almost every data has a 3 dimensional form, 2 dimensions expressing the geographical source of data and 1 dimension expressing the value of data. This is what we call spatial data. Whitener and Davis (1998) state that nearly all business data have a geographic component. They quote an estimate that 85% of all databases contain some form of geographic information, which almost all can be summarized and visualized on a map. These include street addresses, city or state names, ZIP Codes, or telephone numbers with area codes.

For the sake of efficiency, data is conventionally shown as a set of tables and the source of data is expressed as a name substituting its true geographical coordinates, thus limiting it to a 2 dimensional form. When reading this data our brain decodes this name and substitutes it back to its original form. This process lengthens the time needed to comprehend the data and the amount of comprehensible data is limited by the capability of the mind to visualize abstract symbols. Comprehension of spatial data will increase if they were visualized in their original form.

5. Visualization of Spatial Data

Amaripuja (2004) compared the time needed to conduct spatial data analysis with tables, graphs, and maps. In general, maps enable the fastest task execution followed by graphs and tables. The increase in speed offered by maps compared with graphs and tables is most apparent in the Geographical Pattern analysis. In this analysis, the use of 2D maps on average increases speed by 218% compared to tables, and 189% compared to graphs. Amaripuja (2004) concludes that in spatial data analysis, maps should be the main choice of visualization.

Tambe (2001) explains spatial knowledge by quoting the theories of Piaget. From a psychological point of view, Piaget divides the developmental sequence of spatial knowledge into three stages or elements acquired over time: 1) *Landmark Knowledge*, unique patterns of perceptual events that identify a place; 2) *Route Kowledge*, sensory motor routines that connect ordered sequences of landmarks with little or no metric spatial knowledge; and 3) *Survey Knowledge*, two-dimensional layout knowledge of simultaneous interrelations of locations; allows detouring, shortcutting, and creative navigation.

The stages mentioned by Piaget should be kept intact to achieve the maximum benefits of spatial data visualization. Carpendale, Cowperthwaite, and Fracchia (2002) propose a unique technique of visualizing spatial data in order to achieve a high sense of intuition. This technique is incorporated in the three-dimensional pliable surface (3DPS) to create a 3D multi-scale viewing environment for two-dimensional visual information. In 3D the 2D information can be thought of as lying on a planar surface. They further explain that "This surface is manipulated with a simple mathematical function (gaussian curve) allowing for arbitrarily shaped multiple foci and control of the extent and pattern of magnification and compression. This in itself does not make a convincing 3D space. Visual information about the form must be provided that will reveal the nature of the distortion even through areas of sparse data". Carpendale et.al. state that the user presently has the option of applying either the use of shading or gridding. Both of these methods indicate the use of low-level visual skills, but ideally these visual cues should not interfere with the interpretation of the actual information, or create a significant drain on cognitive processing.

The visualization and analysis of spatial data could be aided by the use of a computerized Geographic Information System (GIS). GIS is a computer based system developed for the management and maintenance of spatial data, and integrates data from various time, scale and format. The use of GIS offers managers the ability to obtain and analyse spatial data efficiently and flexibly.

Unfortunately, the generic form of GIS is limited in its quantitative analysis abilities. This problem was acknowledged by Malerba, Esposito, Lanza, Lisi and Appice (2002) by stating that information given in topographic map captions or in GIS models is often insufficient to recognize interesting geographical patterns. Although some prototypes of GIS have already been extended with a knowledge base and some reasoning capabilities, but acquisition of the necessary knowledge is still a demanding task for GIS. They propose a prototypical GIS called INGENS which integrates machine learning tools to assist users in the task of topographic map interpretation. INGENS can be trained to learn operational definitions of geographical objects that are not explicitly modelled in the database.

6. Various Applications of Spatial Data Visualization in Knowledge Creation

Spatial Data Visualization has a bright prospect for increasing intuition and creativity in Knowledge Management. This technique could be used in various applications such as data mining, events detection, geobrowsing, zoomable user interfaces, and spatial data repositories. These applications will be further discussed in this section.

Data Mining. Ben Shneiderman (2002) divides researchers into two separate schools of thought: information visualization researchers and data mining researchers. The first emphasizes on giving insight into the data distributions, while the later utilizes statistical algorithms as the main tool to identify patterns. Shneiderman suggests the use of a combined approach, and recommends that: 1) data mining and information visualization should be integrated to invent discovery tools, 2) users should be allowed to specify what they are seeking and what they find interesting, 3) the social context of users should be recognized, and 4) human responsibility should be respected when designing discovery tools. Shneiderman comes to these conclusions by providing instances where statistical data alone cannot present an optimal solution. He further argues that the emergence of computers allows researchers to analyze large volumes of data without losing data due to statistical biases. Another notion he challenges is the use of developing hypothesis prior to the actual research, which he argues will hinder the ability to seek new possibilities.

Geobrowsing. Various types of data are commonly converted into graphical form. This translation should be carefully done to ensure that they add meaning. On the other hand, geographic data do not have to be translated because graphical data are inherent in them. Unfortunately, this type of data is not used to their upmost potential. Peuquet and Kraak (2002) propose the term “geobrowsing” to designate the process of designing visualization capabilities to facilitate creative thinking for discovering new information from large databases. The authors evaluate the possibilities of geobrowsing by providing several types of maps to represent Napoleon’s Russian campaign. By examining these maps researchers are expected to get a feel of multidimensional geographic data. Gick and Holyoak’s experiment (1980, vide Peuquet and Kraak, 2002) is cited to give statistic proof of a 10 – 80% increase in problem solving abilities due to the use of geographic data.

Zoomable Data Interfaces. The use of the WIMP, the commonly used GUI, has been widespread. Nevertheless, it has its limitations. Unlike the natural environment we live in, WIMPs are static interfaces. Tambe (2001) states that Zooming User Interfaces (ZUI) are becoming increasingly popular as it incorporates an effective and efficient viewing method of larger information spaces on a smaller viewing space. He further explains that “ZUIs are based on the premise that navigation in an information space is best supported by tapping into our natural spatial and geographic way of thinking. The information space in a ZUI is represented by an infinite two-dimensional plane. It allows for scaling of an object on the workspace such that different levels of viewing reveal a greater or lesser magnitude of detail depending on the scale.” In effect, ZUIs has the ability of a zoom lens camera to drill down on an interesting object.

Spatial Data Repositories. The concept of a library usually refers to a repository of books and other printed materials. The advancements of computers and the internet has expanded this notion and we can now easily access digital libraries. However, the main principle remains the same: libraries store information with text related subjects. Crane, Smith and Wulfman (2001) challenge this notion by developing a library which is not only based on text but could be referred to non-textual materials such as maps. They contrast their model for how digital libraries, especially in the field of humanities, are best used with the basic assumptions of many academic digital libraries on the one hand and more literary hypertexts on the other. Crane et.al. developed their system by: 1) generating links from text to text, such as tagging classical languages in order to link inflected words to grammatical analyses, dictionaries and other linguistic support tools, and 2) generating links to visualize time and space, by tagging the names of people, places, and topics, and linking them with reference works that provide glosses and further information.

7. Conclusion

Knowledge sharing discovery, and creation will only be achieved after the transfer of tacit knowledge into explicit knowledge. The ideal medium of transferring tacit knowledge into explicit knowledge would be a medium that can provide a "mental model" that accurately represents tacit knowledge (Showa, 1994). This medium of transfer could either be classification, clustering, visualization or regression (Fayyad et.al., 1966).

Visualizations could be constructed in several forms, such as diagrams, graphics and maps. Considering the fact that 85% of all databases contain some form of geographic data (Whitener and Davis, 1998), most visualizations would be best done by means of Spatial Visualization. This is confirmed by a study made by Amaripuja (2004) which shows that maps are superior to graphs, and graphs are superior to tables. The correct use of spatial data visualizations is thus expected to enhance intuition and creativity in knowledge creation.

References

Amaripuja, P. (2004). *Exploring Differences between Visualization Methods in Spatial Data Analysis*, unpublished, Thesis for Master of Information Technology of Monash University, Australia.

Carpendale, M.S.T., Cowperthwaite, D.J., Fracchia, F.D. (2002). *A Spatial Metaphor for Data Visualization*, from the Web: http://pages.cpsc.ucalgary.ca/~sheelagh/personal/comp/comp96.html

Cheah, C. (2002). *Knowledge Management Lectures Notes*, Monash University, Melbourne.

Crane, G., Smith, D.A., & Wulfman, C.E. (2001). *Building a Hypertextual Digital Library in the Humanities: A Case Study on London*, JCDL'01, June 24-28, Roanoke, Virginia,

Fayyad, U., Piatetsky–Shapiro, G., and Smyth, P. (1996). *Knowledge Discovery and Data Mining: Towards a Unifying Framework*, Proceeding of the Second International Conference on Knowledge Discovery & Data Mining (KDD-96), AAAI Press, Menlo Park, California.

Kennedy, S. (1994). *Unclassed Choropleth Maps Revisited: Some Guidelines for the Construction of Unclassed and Classed Choropleth Maps*, Cartographica Vol. 31, No. 1.

Leung, Y. (1997). *Intelligent Spatial Decision Support Systems*, Springer-Verlag, Berlin.

Malerba, D., Esposito, F., Lanza, A., Lisi, F.A., & , Appice, A. (2002). *Empowering a GIS with inductive learning capabilities: the case of INGENS**, Computers, Environment and Urban Systems, Elsevier Science.

Nonaka and Takeuchi (1995). *The Knowledge Creating Company*, Oxford University Press.

Peuquet, D.J. & Kraak, M.J. (2002). *Geobrowsing: Creative Thinking and Knowledge Discovery Using Geographic Visualization*, Information Visualization, Vol.1.

Shneiderman, B. (2002). *Inventing Discovery Tools: Combining Information Visualization with Data Mining*, Information Visualization, Vol.1.

Smith, D.A. (2002). *Detecting and Browsing Events in Unstructured Text*, SIGIR'02, August 11-15, Tampere, Finland.

Tambe, R. (2001). *The Use of Zooming in Small Screen Interfaces*, research project, unpublished, University of Toledo, December, from the Web: http://jupiter.eecs.utoledo.edu/rashmi/research/need_for_zuis.html)

Vekiri, I. (2002). *What is the value of graphical displays in learning?*, Educational Psychology Review [H.W. Wilson - EDUC]; September.

Whitener, A. & Davis, J. (1998). *Minding Your Business with MapInfo*, OnWord Press, Santa Fe.