Data, Information and Knowledge Transformation

DANA KLIMEŠOVÁ
Czech University of Life Sciences, Prague
Faculty of Economics and Management, Dept. of Information Engineering
Kamýcká 129, 165 21 Praha 6 – Suchdol,
CZECH REPUBLIC
klimesova@pef.czu.cz
and
Czech Academy of Sciences
Institute of Information Theory and Automation, Dept. of Image Processing
Pod vodárenskou věží 4, 182 00 Prague 8,
CZECH REPUBLIC

Abstract: - The paper deals with the relations between data, information, knowledge and knowledge management on the background of the process understanding and the context applying. The contribution of Computer science, Artificial intelligence and the new information technologies is discussed how they can help us to carry out the new knowledge management strategies. The Internet technology and Web services architecture makes data acquiring more open and interoperable in using, offers wide range of context information, has a unique ability to integrate diverse data through shared location and offers real potential for meeting the demands of users. The development and the use of permanently changing information structures open new ways of information processing and understanding. On the example of GIS the context utilisation is presented as a unique approach to the problem of conflict resolution among multiple heterogeneous data sources. The paper discuss the problem of wide context as a tool to compensate and to decrease the uncertainty of data, classification and analytical process at all process to increase the information value of decision support.

Key-Words: - Data, information structures, contextual space, information management, contextual modelling, knowledge

1 Introduction

1.1 From Data to Knowledge

Data can be explained as the product of research or the raw material of information. A single piece of data has no meaning unless the context is understood.

Data needs to be transformed to information that is a flow of messages. The relationship in the data is pointed out and discussed.

Machlup [17] see information is a flow of messages of meanings that might add to, restructure of change knowledge. Dretske [7] offers more useful definitions. He said that, "Information is that commodity capable of yielding knowledge, and what information a signal carries is what we can learn from it."

Knowledge as a meaningful resource has changed the society and economy. The nature of knowledge in contemporary society is so specific that special procedures of control and management have to be used. In the area of knowledge management and knowledge engineering basic terms of these disciplines are data, information, knowledge and knowledge transformation [1].

Knowledge is a multifaceted concept with multilayered meaning. The traditional epistemology adopts a definition of knowledge as "Justified True Belief". In theory of knowledge creation, knowledge is seen as a dynamic human process of justifying personal beliefs as part of an aspiration for the "truth". Knowledge is often not explicitly describable not easy to explain and to formulate and formalize.

Therefore the knowledge is recognized as *explicit* and *tacit.* Knowledge creation is a spiralling process of interactions between explicit knowledge [21], [18]. The interactions between the explicit and tacit knowledge lead to the creation of new knowledge. The combination of these two categories makes possible to conceptualise four conversion steps: Socialisation, Externalisation, Combination and Internalisation.

Another well-known distinction is between *declarative knowledge* (object knowledge) and *procedural knowledge* (or process knowledge). In the literature is also used *to know what* and *to know how*.

Consider the observation made by Neil Fleming [10] as a basis for thought relating to the following diagram.

- A collection of data is not information.
- A collection of information is not knowledge.
- A collection of knowledge is not wisdom.
- A collection of wisdom is not truth.

The idea is that information, knowledge, and wisdom are more than simply collections. Rather, the whole represents more than the sum of its parts and has a synergy of its own.

1.2 Knowledge Management

Knowledge management recognizes the need to exploit intellectual capital, but many practices fall short by only concentrating on individual knowledge components. Integrated knowledge has multiform structure, appropriate links with possibility to aggregate, separate and generalize parts and wholes in temporal dependence.

Moreover, knowledge management needs the tool that can help to supply users by relevant and suitable knowledge in appropriate time and in consumable integrated form when they solve tasks like process improvement, decision support, training, risk management and others. Mentioned tool would be effective used in future to auto-learning enhancement and integration of knowledge base instead of storing and accessing information from a central repository.

However, the specific introduction of the term knowledge creates a different viewpoint and allows more sophisticated ways of the use. It is no longer sufficient to deliver huge amounts of information to users instead it is important to support them in doing their knowledge work.

Knowledge management is a holistic approach, which can be analyzed in different ways. For this reason, it is difficult to give an exact definition. Nonaka & Takeuchi, [21] stress in their definition the importance of the distinction between tacit and explicit knowledge: *Knowledge Management is the tacit and explicit knowledge framework for a dynamic human process of justifying personal belief toward the truth.*

Explicit knowledge is knowledge that is already extracted and consumable in books or other media. Tacit knowledge is not present in explicit form, and can often not even be articulated by a person who possesses the knowledge.

Thomas Davenport [6] stresses the importance of the process and supply chain:
Knowledge Management is a formal, structured initiative to improve the creation, distribution, or use of knowledge in an organization. It is a formal process of turning corporate knowledge into corporate value. Karl Erik Sveiby also emphasizes the corporate value [3], [26]:
Knowledge Management is the art of making money out of immaterial assets.

We are mainly interested how the knowledge can be captured and processed, what technology areas can help us to realize a *Knowledge management strategy*, and what kind of knowledge is managed in fact.

1.3 Information System

It is necessary to understand the role of information systems, and especially how the evolution of the Internet and Intranets can contribute to effective knowledge management. These themes shape the role of the technological infrastructure in knowledge management. The purpose of an information system is to answer the questions. Information must be structured (knowledge must be represented) so that answers can be found among the information in the system or derived from the information in the system.

To arrive at a better answer, we need a still more complex system with still more facts and rules. Further data to take into account include more information about the interactions, possible incompatibilities, use experiences, effectiveness and may be costs. A system containing all these data for a large number of objects can answer more questions and would be called an expert system.

Main characteristics of a good information system are as follows:

- The simple way of communication with user that allows the interpretation of requests and incorporation of user confirmation when needed (additional conditions).
- Adaptive ability is requested to the special needs of the user and the specific situation.
- The system is able to ask for more information if it is needed to derive a good answer.
- System gives answers in easily-understood format.
- Gives reasons for suggested problem solutions, explains its reasoning.
- System assists in knowledge acquisition and is able to learn,
- Accepts different information structures,
- Has in disposal a wide range of information processing techniques and methods,
- And appropriate knowledge presentation tools (interpreting language, graphical representation, generating language and graphics).

The evolution of the Internet and Intranets can contribute to effective knowledge management. And further, these two main strands of current management focus - the Internet and knowledge management, in combination, they provide a powerful driving force for business and individual opportunities.

2 Information arrangement

Web sites are built around the basic structural themes. These fundamental architectures govern the navigational interface of the Web site and they are simple understandable for users. Three essential information structures can be used to build a Web site: sequences, hierarchies and webs.

2.1 Sequences

The simplest way to organize information is to place it in a sequence. Sequential ordering may be chronological, a logical series of topics progressing from the general to the specific, or alphabetical, as in indexes, and encyclopedias.

Straight sequences are the most appropriate organization for training sites, for example, in which the reader is expected to go through a fixed set of material and the only links are those that support the linear navigation path. The sequence is simple and very good predicable linear arrangement, see Fig.1.

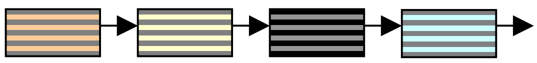

Fig. 1 Linear access

More complex Web sites may still be organized as a logical sequence, but each page in the main sequence may have links to one or more pages of additional information, parenthetical information, or information on other Web sites – Fig.2.

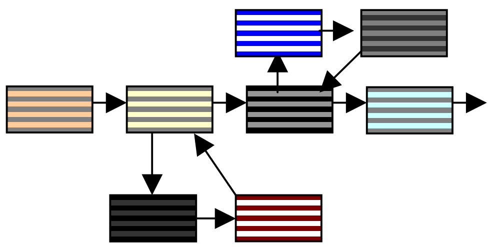

Fig. 2 Logical sequence

2.2 Hierarchies

Information hierarchies are the best way to organize most complex bodies of information. Because Web sites are usually organized around a single home page, hierarchical schemes are particularly suited to Web site organization. Hierarchical diagrams are very suitable to incorporate own point of view on system arrangement and most users find this structure easy to understand. A hierarchical

ISSN: 1790-5117

ISBN: 978-960-474-064-2

organization allows to simplify complex package of information and to apply analytical approach to content organisation.

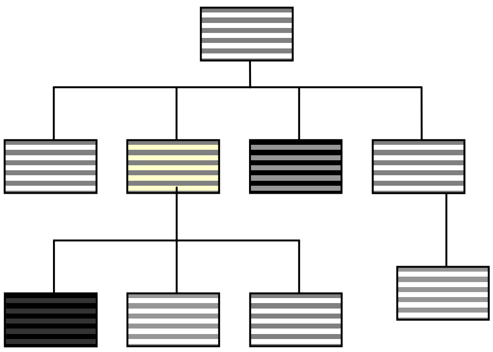

Fig. 3 Hierarchical arrangement

2.3 Webs

Web organizational structures contain only few restrictions on the pattern of information use. In this structure the goal is to follow associative thought and the free flow of ideas, allowing users to keep their interests in a unique, heuristic pattern investigation. This organizational pattern develops the possibility to follow complicated links both to information elsewhere in the site and to information at other sites.

Although the goal of this organization is to exploit the Web's power of linkage and association to the fullest this seemingly simple structures can propagate confusion. Ironically, associative organizational schemes are often very impractical structure because they are so hard to be understood and predicted by user.

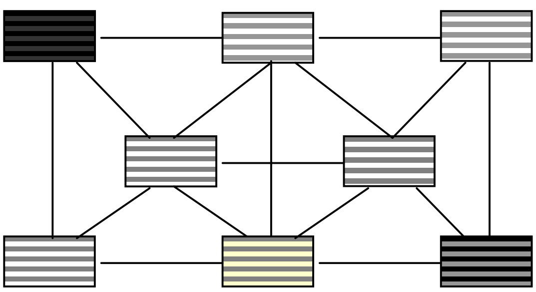

Fig. 4 Cobweb arrangement

Webs work best for small sites dominated by lists of links and for sites aimed at highly educated or experienced users looking for further education or enrichment and not for a basic understanding of a topic.

3 Contextual Space

The fast development of GIS is closely connected with very suitable information structure – dynamic structure with conditional access that is convenient for the multi-source, multi-criteria and multilayer way of the work with wide context consideration.

3.1 From Information Management to Knowledge Management

Knowledge management is considered as the next source of competitive advantage. The computers supports knowledge processes from processing and collecting knowledge to dissemination and sophisticate use.

The development of the Internet and related technologies supports communication. It uses widely supported communication standard protocol and it is accessible from many locations and through many different computer platforms. The communication plays basic role in the information lifecycle as well as dissemination. The ITC technology can from the point of knowledge management support further important features like: collaboration, connection and conversation when we are dealing with tacit and explicit knowledge.

The area in Computer Science that is most influenced by the concept of knowledge is Artificial Intelligence and Knowledge Based Systems [6].

In the early 1980s the development of a knowledge based systems were seen as a process of transferring human knowledge to an implemented knowledge base. This transfer was based on the assumption that the knowledge, which is required by the system already exists and only has to be collected and implemented [5].

Since the knowledge is specified independently from the application domain, reuse of the knowledge is enabled for different domains and applications.

The knowledge modelling connected with knowledge based systems is influenced everyday by new research results.

We have to consider that the model is only an approximation of reality and the modelling process is a cyclic process. New observations may lead to a refinement, modification, or completion of the already constructed model.

On the other hand, the model may guide further acquisition of knowledge – contextual understanding and moreover, the modelling process is dependent on the subjective interpretation of the knowledge.

3.2 Context Understanding

The contextual modelling deals with different types of context information. It is possible consider context as the reflection of object or phenomena using different interpretation through the system of cognition: *perception, conception, interpretation.*

We can understand the context as the reflection of selected facts concerned with validity of statements and the system of argumentation: *identification, analysis* and *synthesis.*

Context has to be catch as the reflection when hypothesis stays instead of experience in the system of instinct based context: *recognition of patterns, coordination by intuition, judgement due to synthetic inference.*

The context can be considered as the reflection concerning validity of statement using knowledge generating system: *analogy to experience, deduction from model of ideal world, induction.*

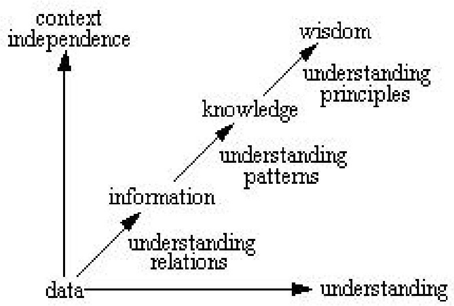

Fig. 5 Shows the relationship between context and understanding and transformation process from data to knowledge [30].

Applying context it is possible to derive new quality of information that can be used to support decisions, fig. 5 and fig.6. While information reflects an understanding of the relations between data, it generally does not provide a foundation for why the data is what it is and how the data is changed over time. Information has a tendency to be relatively static in time and linear in nature.

Information is a relationship between data with great dependence on context for its meaning. Information relates to description, definition, or perspective (what, who, when, where).

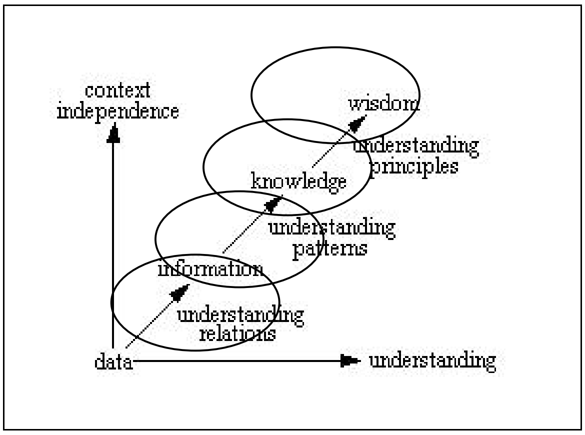

Fig. 6 The transformation processes are more and more sophisticated.

Instead of knowledge that contains the method or approach (how to do things), include practice and strategy. Thinking about upper mentioned facts what type of structure would be appropriate to follow the context? It seems that to fulfil drawn imagination we have to consider dynamic structure where the *dynamic* means that the move will be inherent part of that structure. It could be some structure derived from Brownian web, fig. 7.

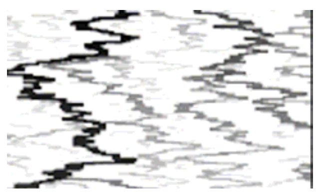

Fig. 7 Brownian web [27]

Brownian paths are motions on Brownian web and just Brownian motion is the key point of modern probability theory that would be able to describe the measure of applied context and the resulting direction of the transformation process.

3.3 Uncertainty Management

Very heterogeneous data available in the combination with multi-criteria conditions evaluation causes increasing of uncertainty we meet in the processes and consequently in results. The dimensionality of data and the complexity of objects structure hierarchy are rapidly growing and consequently with these aspects increase the uncertainty entering into the processing.

Data uncertainty plays a special role in the environment of Internet and Web Services [3] in connection with GIS. It is quite another situation than in case of the closed system, where the user has full control over all steps of processing from data input to presentation of results. In frame of open interoperable system with access to web sources with a great number of existing databases the user control gets completely lost.

A great number of existing databases offer a variety of data sets covering different thematic aspects like topographic information, cadastral data, statistical data, digital maps, aerial and satellite images including temporal data. Data collection is changing from digitising own data to retrieving and transferring from existing databases coming from task processing and result presentation.

To deal with such data sets, the user requires an uncertainty description that has to be added by the producer. User needs an appropriate uncertainty model for this purpose, integrated in GIS [4].

From the philosophical point of view the uncertainty is quite natural part of our life and the surrounding world. Uncertainty might not be a bad thing if you can make better use of it then others, in the sense of competition. We can understand this concept in the frame of description - ordered and chaotic.

Usually we meet uncertainty in the sense of valuation. Uncertainty is a real and universal phenomenon in valuation and the sources of uncertainty are rational and can be identified. Valuation is the process of estimating the value and estimation will be affected by uncertainties. The input uncertainties will translate into an uncertainty of the valuation.

Actually – the uncertainty arises from imperfect understanding of the events and processes in the world around. From another point of view the fact of uncertainty is very stimulating for the research on the field of *defining*, *measuring*, *modelling* and *visualizing* uncertainty and data quality analysis. The uncertainty opens the space for further questions like: *where, why and when* and the answers to this question can help us to do better decision [11].

To gain the relevant answer it is necessary incorporate the various contexts into the analysis of objects, phenomena, events and processes and connect up uncertainty into the knowledge-construction and decision-making process through context cognition.

Open systems are using frequently various models but the model is only an approximation of reality and the modelling process is dependent on the subjective interpretation of the knowledge. It means that new observations may lead to a further refinement, modification, or completion of the already constructed model. And the model may guide further acquisition of knowledge and the knowledge is the base for decision support. Moreover, besides knowledge modelling also knowledge representation is very important field of research.

Data are not perfect from many reasons:

- Incomplete data
- Precision of measurements
- Discreet description of connective phenomena
- Inherent part reflecting our understanding of things [23]

On the other hand the current top level of GIS usage, it is control GIS, where the large ability is aided to implement knowledge models from different branches of scientific investigation, wide context implementation including less evident connections, models of trends, objects and expected or predicted relations.

To reduce uncertainty of data it is mainly the question of the proof of recognized quality assurance. Some users often take the pragmatic approach to the cost versus accuracy. Sometimes, without the relevance testing, the resolution of data is used for the whole set of different task. Then the problem of over-defined and under-defined objects brings the difficulties [14].

Especially uncertainty of a geographic object can be modelled through uncertainty of its geospatial, temporal and thematic attributes. Uncertainty of relations takes into consideration spatial, temporal and spatial-temporal relations.

To add suitable attribute or to spread the net of relations reduce the uncertainty of the object. The special case is to model objects uncertainty using spatial-temporal approach to the objects and incorporate spatial-temporal relationships. The dynamics of object is very powerful tool to obtain exact results about the object and phenomenon behaviour to support further decision [28].

The decision making process is always associated with some level of uncertainty which can rise from:

- Definition of the problem
- Data used
- Sequence of operations used to obtain result
- Understanding of result

GIS is shifting very fast from desktop GIS to network GIS. Great advantage of network GIS is ability to provide GIS services in a networked environment, typically through the Internet.

With this technology, all GIS components, data components and functional objects, can be distributed across the network. In this component-oriented framework the user has no problem with the increasing complexity of information structures and quality demands and is able understand objects and phenomena and theirs expressions in various contexts and provide richer analysis with different aspects of modelling.

The integral part of control GIS [15] is the modelling where the information layers from real, artificial and virtual world are composed together to select optimal scenario or verify given hypothesis or assumptions.

The contextual design of spatial data and further development of geo-information technologies, image processing techniques and the possibilities of object history modelling together with the geographical networks environment will provide quite new and considerably wider possibilities of using GIS.

GIS architecture is open to incorporate new requirements of knowledge-based analysis and modelling, namely in connection with web designed spatial databases and temporal oriented approaches. This type of geo-information processing it is the resource, tool and means. It is modelling in most common sense.

If the standard geographical database is understand as a digital model of the real world than control GIS handles the DB, which is the result of temporal interface of standard DB with virtual and artificial models of real world.

Temporal or dynamical analysis of spatial data is needed in various fields such as mainly known environmental systems analysis. Dealing with this approach we are facing the difficulties in generating spatial-temporal space of quality data for analysis, the necessity of interpolation or integration of observational data [23].

The great advantage is to mix spatial topological relations with temporal topological relations and generate and extract new spatial-temporal relationships from the spatial-temporal objects. The uncertainty of objects is projected to the uncertainty of relationships between objects.

Context sensitive object recognition is a successful strategy to reduce uncertainty geographical objects and primarily it holds in temporal context application.

On the base of these considerations it is possible to create temporal space, to define objects type and propose the concept of interior and border of the interval and than the algebra can be built. It is of course far from our exciting imagination to build real temporal database.

3 Conclusion

The contribution deals with more abstract level for reflection and understanding of the various modelling processes. In this paper, the problem of wide context is discussed. Our decisions are becoming increasingly dependent on understanding of complex relations and phenomena in the world around and context modelling is able to incorporate new requirements and produce more valuable results. The main goal has been to show selected aspects of this process and compare the increasing possibilities of the sources with the difficulties of data contextual structuring and the object dynamics implementation.

Acknowledgements
The Project Information and knowledge support of strategic control - MSM 6046070904 supports the work. This support is very gratefully acknowledged.

References:

[1] Aamodt, A. and Nygard, M., 1995. Different roles and mutual dependencies of data, information and knowledge. Data & Knowledge Engineering, 16, 191-222.
[2] Benedikt J., Reinberg S., Riedl L., 2002. A GIS application to enhance cell-based information modeling. Information Sciences 142 (2002): 151-160.
[3] Bernbom, G., 2001. Information Alchemy: The Art and Science of Knowledge Management, EDUCAUSE Leadership Series #3. San Francisco: Jossey-Bass. Graham, Ricci.
[4] Bolloju N., 1996. Formalization of qualitative models using fuzzy logic. Decision support systems 17(1996): 275-289.
[5] Cornelis, B., and Brunet, S., 2000. "A policy-maker point of view on uncertainties in spatial decisions". "Spatial data quality" (Shi W., Fisher P., and Goodchild M., Eds), Chapter 12, pp. 168-185.
[6] Davenport, T. H. and Prusak, L.: *Working Know ledge. How Organizations manage what they know*. McGraw-Hill; Harvard Business School Press, 1998.
[7] Dretske, F. I.: *Knowledge and the flow of information*. Basil Blackwell Publisher, 1981.
[8] Erdmann, M.: *Ontologien zur konzeptuellen Modellierung der Semantik von XML*. Dissertation, Institut AIFB, University of Karlsruhe, 2001
[9] Fensel, D., Decker, S., Erdmann, M., and Studer, R., 1998. Ontobroker: Transforming the WWW into a Knowledge Base. In Proceedings of the 11th Workshop on Knowledge Acquisition Modeling and Management, Banff, Canada, April 18-23.
[10] Fleming, N. Coping with a Revolution: Will the Internet Change Learning?, Lincoln University, Canterbury, New Zealand, 1997.
[11] Fuller R., 2000. In: Introduction to Neuro-Fuzzy systems. Advances in soft computing, Physica-Verlag Heidelberg. 289 pages.
[12] Heflin, J.: Towards the Semantic *Web: Knowledge Representation in a Dynamic,* Distributed Environment. Ph.D. Thesis, University of Mary land, College Park, 2001.
[13] Hlupic, V., Pouloudi, A. and Rzevski, G.,"Towards an integrated approach to knowledge management: 'hard', 'soft', and 'abstract' issues.", *Knowledge and Process Management*, 2002, **9**(2), 90-102.
[14] Klimešová D., 2006. Study on Geo-information Modelling, 5 (2006), WSEAS Transaction on Systems, pp. 1108-1114.
[15] Klimešová D., 2004. Geo-information management. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 35 (2004), 1, pp. 101-106.
[16] Lynch P.J., Horton S., Web Style Guide: Basic Design Principles for Creating Web Sites, New York, 2002.
[17] Machlup, F. and Mansfield, U. (Eds.), *The study of information: Interdisciplinary messages* (pp. 3–59). New York: John Wiley, 1983.
[18] Martensson, M. (2000). "A critical review of knowledge management as a management tool." *Journal of Knowledge Management*, **4**, 204-216. Milton N. R.*Knowledge Technologies*. Milan: Polimetrica, 2008.
[19] Milton, N. R. *Knowledge Acquisition in Practice: A Step-by-step Guide*. London: Springer, 2007.
[20] Milton, N., Clarke, D. and Shadbolt, N. (2006). Knowledge Engineering and Psychology: Towards a closer relationship. *International Journal of*

Human-Computer Studies, Volume 64(12), 1214-1229.
[21] Nonaka, I., Takeuchi, H.: *The Knowledge Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, 1995.
[22] O'Leary, D. 1998. Knowledge Management Systems: Converting and Connecting. IEEE Intelligent Systems, May/June 1998, pp. 30-33.
[23] Parent C., Spaccapietra S., and Zimanyi E.,2000. MurMur: database management of multiple representations. Proceedings of AAAI-2000 Workshop on Spatial and Temporal Granularity, Austin, Texas.
[24] Peuqeut, D.J., 2002. Representations of Space and Time. The Guilford Press.
[25] Power, D., J., 2002. Decision Support Systems: Concepts and Resources for Managers, Quorum Books Published 2002.
[26] Sveiby, K. E. and Lloyd, T.: *Das Management des Know-how*, Frankfurt am Main, Campus Verlag 1990.
[27] Swart J.M., The contact process seen from a typical infected site, *Journal of Theoretical Probability*, 2008.
[28] Zhang J., Goodchild M. 2002. Uncertainty in geographical information. Taylor & Francis, London, pp. 127-130.
[29] <http://www.cyberartsweb.org/cpace/ht/thonglipfei/ct_dfn1.html>.
[30] http://www.outsights.com/systems/intst/int.htm, Systems Thinking, The Knowledge Centered Organization, United States, Bellinger G., 2004.