

Collaborative Platform Using Knowledge Cartography - ToxNuc-E

Reena T. N. Shetty¹, Joël Quinqueton², Pierre-Michel Riccio³, Jean Michel Penalva³, Jean Villerd³

¹Laboratory EMA/LGI2P, ENSMP, Paris

reena.shetty@ema.fr

²Laboratory LIRMM, UM II, Montpellier

jq@lirmm.fr

³Laboratory LGI2P, EMA, Nimes

pierre-michel.riccio@ema.fr, jean-michel.penalva@ema.fr, jean.villerd@ema.fr

ABSTRACT

Our research revolves around a platform called ToxNuc-E entirely dedicated to the research activities in studying effects of nuclear toxicology on plants and animals. Here, researchers from different domains interact and exchange information using our platform as the common ground involving new concepts, methods and services to encourage collaborative work for their research activities. The work is based on the co-operation and collaboration between biologist, chemists, sociology specialists and engineers building the platform. In our proposal we attempt to capture the individual information and then share it with every other member of the platform to build a group representation of knowledge, attracting and encouraging people to participate in collaborative tasks.

KEYWORDS: Frameworks and Methodologies for Collaboration, Tools for Collaborative Environments, Coordination, Cooperation and Collaboration, Collaboration Enabling Technologies, Platforms for Collaboration, Collaborative Knowledge Management.

1. INTRODUCTION

Ages ago, collaborative or group oriented work started as a necessity when people could not attain their objectives individually. Personal interests and desire for gain lead humans and animals to form groups, boosting productivity and innovation. In the current world of information, collaboration is no longer limited to living beings and has been consequently extended to machines, considering their growing importance. There has been a revolution in the domain of collaboration with the introduction of various applications and tools making collaborative work efficient, faster and well organized. Thus in turn promoting concurrent advantages for companies and organizations involved.

Collaboration literally means an action or a work completed in common with two or several persons. It is a group activity where individuals unite to form alliance or unions with an intention to attain an objective. We can find evidence of group activity in many living beings ensuring early completion of tasks and better security against possible dangers. Every member in the group experiences better results when tasks are being accomplished in a coordinated group than attaining it individually.

This meaning of collaboration as a form of collective intelligence has been stated by J. M. Penalva et al [1] as a hypothesis relative to the capacity of a group of cognitive actors and artificial agents to reach a higher level of performance than the addition of individual ones. Based on this definition, collaboration appears as a form of coordination [2] which is managing dependencies between activities in a group.

The study of mechanisms of collaboration and sharing is relatively a developing area in research, proposing new methodologies for the capture and re-use of organizational knowledge. Successful application of collaboration practices involves the understanding and constructive use of organizational learning and information flows within the organization. The concept of collaboration is becoming more important in the evolving context of global network, thus placing the user at the centre of a collective device. Collaborative work can either be of the nature, where each group member is involved in every activity with the work being highly interactive or where each group member is given an individual task.

In our proposal, we start by building a user friendly platform with built-in applications and features to guide the users in efficiently using the platform for information retrieval and management and also encourage them to share information about their research activities with the

rest of the registered members belonging to different research communities, on the platform. In this paper we mainly focus on the utilization of knowledge cartography which acts as a mirror of the group activity carried out on the platform both in terms of research teams and members involved.

The following sections of this paper are organized as follows: In section 2 we give an introduction about the creation and use of our application platform ToxNuc-E followed by section 3 where we describe our work on knowledge cartography starting with Bubble of Confidence among researchers applicable on any collaborative platform. In section 4, we propose novel approaches for information retrieval and management to boost collaborative thinking in group oriented working environment. This is basically a knowledge representation model called the Extended Semantic Network (ESN).

The subsequent section explains how the results of ESN are used by the application Molage on our platform. Finally, in the last section we conclude this paper with the summary of the work completed, advantages and results studied and direction for future work.

2. ENVIRONMENTAL NUCLEAR TOXICOLOGY PLATFORM (ToxNuc-E)

In 2001, a multi-field inciting program was initiated by the CEA (Commissariat à l'Energie Atomique in France) to stimulate the emergence of a community of experts and young researchers around a stake, mainly touching the public health and the environment. It mainly handled the question of understanding the mechanisms of actions of heavy metals and radio nuclides on the various levels of organization of the living beings. This research program in fundamental extended in 2004 to four organization of research partners (CEA, CNRS, Inra and Inserm) involving some of the well established French research laboratories is a multi-field project which involved a great number of researchers from different research disciplines like biology, chemistry, medicine and physics.

The main tasks of the direction of this program were to manage and provide all the necessary tools and applications for easy interaction among the vast community of researchers involved in the program. Thus at length favoring and supporting communication leading to information exchange between actors (researchers) of the program: grounded on a collaborative platform.

The idea of this platform rose from several questions on nuclear toxicology like: What are the effects on the living

organisms from the elements such as the radio nuclides or heavy metals and metalloids used in medicine, research or for industrial activities? How a toxic element does reach its molecular target? Why certain cells of a body are more sensitive or state? How certain cells or organizations resist these elements which can be toxic?

Answering these questions eventually made it possible to have a thorough knowledge of the impact of the anthropical activities on human health and its environment. The recent studies and observations made on impact of toxicology on mankind are very few in France as well as abroad. Some of the field and methodologies used scarcely integrate the projections of the revolutionary techniques proposed by genomic and biotechnology domains.

Contradictory to these observations, research in biology and genetics is developing at a vertiginous speed and all the resources of post-genomic available to renew the field of toxicology are highly neglected in biology. In order to contribute to this society and human health related questions, fresh impulse was given to this research within the framework of an inciting multi-field Program heading now "Environmental Nuclear Toxicology: ToxNuc-E" [3].

2.1. Scientific Objectives of the Program ToxNuc-E

The program mainly focuses on the question of including the mechanisms of actions of heavy metals and radio nuclides on the various levels of organization of living organisms (molecular, cellular, bodies and fabrics, whole organizations) in order to propose preventive technical solutions, provisions of effective monitoring and solutions to decontaminate these elements distributed in certain compartments of the tropic chain.

These chemical elements were primarily identified in dialogue with various actors involved in nuclear die, in industry and in research, and a list of interesting elements were identified and brought out. These elements are listed as follows: tritium, beryllium, boron, carbon, cobalt, selenium, strontium, technetium, cadmium, iodine, cesium, lead, uranium, plutonium, americium, zinc, copper and nickel.

The state of the art on this domain was very weak to be used in identifying these elements considering the fact that very little work been carried out in this field. This resulted as a primary factor in focusing the studies on two fields called environmental toxicology and human toxicology. In these two fields, it is a question of being interested: with the biological effects as of these

substances and the molecular and cellular mechanisms of transport, of toxicity and de-toxicity. This subsequently leads to the issue of proper co-ordination between researchers from the two different fields.

- For environmental toxicology, it helps to study the mechanisms of transfer of the geo-sphere towards the biosphere by means of the bacteria and the plants and to imagine applications to decontaminate the terrestrial or watery environments.
- In human toxicology, it helps to imagine applications for treatment of contamination by targeting the studies on uranium and plutonium. The organizations on which these studies are focused are preferentially those whose genome is sequenced i.e.: bacteria, yeast, arabidopsis, human cells, mouse, and rats. This approach allows the massive use of the methods of genomic.

2.2. Mobilization and Organization of the Program

The human means does exist but it is a question of mobilizing them on some clearly given scientific objectives. The direction of the program organised meetings for including some of the major researchers in the biological, chemical, physical and informative fields. Committees were organised and co-ordinators or heads for each research project were chosen. Several researchers geographically dispersed were brought into contact through this platform.

The registered members were over 700 researchers from diverse fields working on topics related to nuclear toxicology. In a very short period, vast information was collected on the platform. Now the problems like efficient data management, easy information retrieval and safety about sharing one's research results with other members of the platform known only through professional contact because of similar research interests needed to be resolved.

A positive response to this question would automatically encourage researchers on the platform to exchange information and discuss the research requirements and observation with other members of the community. Thus leading to a collaborative proceeding to resolve issues concerning to nuclear toxicology. Precisely, to this requirement using knowledge cartography we basically try to provide a visual image of the collaborative work done on the platform and help them achieve confidence on each other and make co-ordination easy through the utilization of our proposed tools on the platform. Using these tools will actually permit to have a global knowledge about the research work carried on the

platform and as well provide information about the researchers involved and their actual domain of interest. This will in turn boost confidence and encourage a will for collaboration

3. BUBBLE OF CONFIDENCE FOR ToxNuc-E PROGRAM

The concept of bubble of confidence [4] consists in providing the researchers a set of visual representations of the collective group or program in which they work and evolve, in the structuring of their field. It also consists in showing them their position in the respective community. More precisely; the bubble of confidence makes it possible for the researcher to locate, on charts and identify members who are "his closer neighbours" on scientific dimensions which interest him [5]. Studies have shown that, unlike the face to face situations where interaction process is blocked by difference of opinions and points of view, asynchronous computer aided communication allows a suitable communication process, towards collective intelligence

The researchers and the projects constitute the most important entities of the program. In figure 1 we see that the projects and the coordinators of projects are presented as the most important entities which play the part of reference scientists, subsequently allowing the other researchers to position themselves based on these benchmarks.

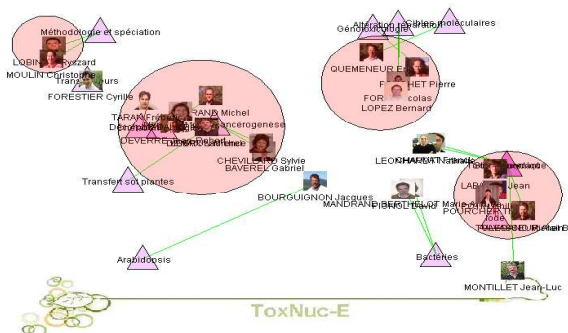


Figure 1. Bubble of Confidence among Researchers of the Program ToxNuc-E

The clusters are the regroupings which are created corresponding with entities which have joint properties, with more or less, strong or higher values. These actually determine the proximities or the distances between the entities. The assumption with which we advance is that the bubble of confidence will have a positive influence on the behaviour of the researcher with respect to the use of the technological device.

This in turn becomes a means of socialization, creation of bonds and interactions, which is favourable for building an actual collective intelligence. The concept is presented in figure 1 where bubbles of confidence are formed around researchers closely located in terms of research activities. This finding in turn creates a professional bonding among researchers encouraging them to share research information with other members of the platform and to collaboratively work to attain the desired goal.

4. EXTENDED SEMANTIC NETWORK FOR INFORMATION RETRIEVAL

Extended Semantic Network [6] is an innovative tool for knowledge representation and ontology construction, which looks for sets of associations between nodes semantically and proximally as opposed to the present method of keyword association. Our objective here is to achieve a semi-supervised knowledge representation technique with good accuracy and minimum human intervention, using the heuristically developed information processing and integration methods. This model is built based on the information and research documents shared on the platform ToxNuc-E.

4.1. Hybrid Approach – Extended Semantic Network (ESN)

The basic idea of Extended Semantic Network is to identify an efficient knowledge representation and ontology construction method to overcome the existing constraints in information retrieval and classification problems on ToxNuc-E platform. To realize this, we put our ideas into practice via a two phase approach. The first phase consists in processing large amount of textual information from the platform using mathematical models to make our proposal of automatic ontology scalable. The second phase consists in examining carefully and efficiently the various possibilities of integrating information obtained from our mathematical model with that of the manually developed mind model.

The first phase of our proposal is carried out by realizing a network of words mathematically computed using different statistical and clustering algorithms. Thus creating a proximal network computationally developed, depending essentially on word proximity in documents. The second phase is ensured by a heuristically developed method of network extension using the outputs from the mathematical approach. This is achieved by considering the manually developed semantic mind model as the entry point of our concept network.

Here, the primary idea is to develop an innovative approach obtained by combining the features of man and machine theory of concepts, whose results can be of enormous use in the latest knowledge representation, classification, retrieval, pattern matching and ontology development research fields. In this paper we discuss and highlight the methods used by us for information processing and integration aimed at visualising a novel method for knowledge representation [7] and ontology construction [8]. This will help the ToxNuc-E researchers to easily retrieve information on the platform and will encourage information sharing.

4.1.1. Proximal Network for Efficient Data Processing

Proximity is the ability of a person or a thing to tell when it is near an object, or when something is near it. This sense keeps us from running into things and also can be used to measure the distance from one object to another object. The simplest proximity calculations can be used to calculate distance between entities thus avoiding a person from things (s)he can hit. Proximity between entities is often believed to favour interactive learning, knowledge creation and innovation.

The basic theory of proximity is concerned with the arrangement or categorisation of entities that relate to one another. When a number of entities are close in proximity a relationship is implied and if entities are logically positioned; they connect to form a structural hierarchy. This concept is largely used in medical fields to describe human anatomy with respect to positioning of organs.

Our Proximal Network Prototype model is built based on this structural hierarchy, of word proximity in documents. This approach is mainly employed to enable processing of large amount of data in a considerably small time. Another important aspect of this approach is its ability to automatically process the input data into a network of concepts interconnected with mathematically [9] established relations.

For building this prototype we systematically employ three phases for identifying our data to build the final network. We first start with a set of documents related to 3 major fields out of the 15 fields in the nuclear environmental toxicology domain, furnished by the program ToxNuc-E. The documents obtained are first converted into simple text format using an external converter and are later input into our first stage called the pre-treatment process. This process is carried out in 2 different approaches, one for identifying the most significant words and the other for eliminating hollow terms. Here, in this process the input document is processed in several stages and an output of word

connecting these concepts. The concept network thus built is based on the meaning each concept pair shares.

They have been currently chosen on an experimental basis, after proper consideration and analyzing the requirements of our approach. We start with our domain name representing the super class in our approach. The super class is then connected to its subclasses based on the category of the relation they share, which can be chosen from the four links we provide.

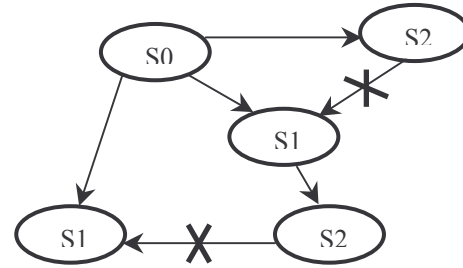


Figure 4. Relational Flow Illustration

The figure 4 illustrates this idea where the relational flow is possible from S0, S1 node towards S2 but not vice versa. Here S0 represents the first level, S1 the second and so on.

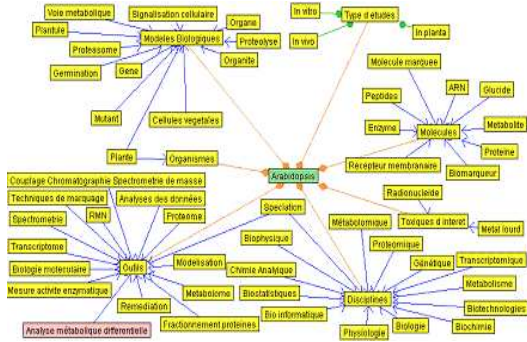


Figure 3. An Extract of Semantic Representation of Concepts for Arabidopsis

The objective here is to introduce the semantic based relation into our mathematically modeled proximal network. The network thus developed is then analyzed and merged to obtain one single semantic network for that domain. This process is repeated on different lists of concepts concerning various domains to obtain one network for each domain. The result thus obtained is input into the Mysql database along with the relational links they share. This is then visualized and edited using the graph editor. The figure 3 shows a semantic network developed on the project arabidopsis.

4.1.3. Integration of Mind and Mathematical Model to Obtain ESN

The 2 different resulting models obtained by our methods dilated in the previous sections are combined employing the simple extension methods. We call this the hair extension method. Here we start with semantic network retaining all its nodes as the starting core of our network. This network is then expatiated by adding nodes from proximal network.

Here we identify the common nodes between the two models and then add on the rest of the node network connected to this node from the proximal network. Presently we have limited this to a level of 5 extensions i.e. only the next 5 level of nodes are added from the proximal network. We also use the process where the relational flow is possible only from a lower level node to the upper level node.

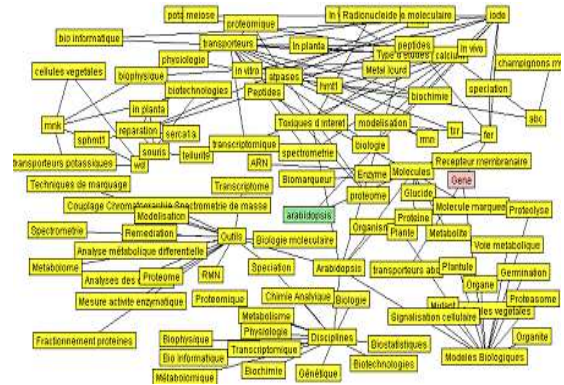


Figure 5. An Extract of Extended Semantic Network Visualised using Graph Editor

Simultaneously, several other optimizing algorithms are being considered to be utilized in merging the networks to build the Extended Semantic Network depicted in figure 5. We are exploring the possibilities of using the genetic algorithms and features of neural networks to enhance these results.

4.2. Results and perspectives of ESN

Our present results have been verified by experts in comparison with human developed ontology and concept networks and validated for providing satisfying results, as stated in the next section. The results of our algorithm have been judged to provide results very close to human constructed concept networks. It has also proved to take much less time for construction and highly cost effective. The ESN prototype thus targets at initialising a new method for knowledge representation for easy ontology construction which can be employed in new generation search algorithm to facilitate information management, retrieval and sharing.

We are on the conclusion that the results are exceedingly encouraging in terms of accuracy. The next step will be to include natural language processing techniques like to our pre-treatment process. Our objective is to develop an application for document classification and indexation based on the results of Extended Semantic Network. This application library is intended to be used for classification purpose in the project ToxNuc-E for better data management on the platform.

Our future work on Extended Semantic Network is to include user modeling features by monitoring the behavior; interests and research works carried out by the members of ToxNuc-E and then build a model unique to each user. This model consecutively builds a profile for each user and sequentially stores the details obtained in a database. These details can be utilized to better understand the user requirements thus helping the user in efficient data search, retrieval, management, and sharing.

Some of the major points we hope to achieve through this method of knowledge representation network are

- To make construction of semantic based concept networks cost effective by campaigning minimum human intervention. In turn reducing the construction time using mathematical models.
- To identify a good balance between mind and mathematical models to develop better knowledge representing networks with good precision and high recall.

4.3. Document Classifier

As an application of our model a document classifier has been developed and integrated on the ToxNuc-E platform. This document classifier uses the ESN knowledge model to classify documents [12] based on their inclination to the 15 research projects that are being piloted by ToxNuc-E. For initial testing we selected a set of new publications from the project related to Arabidopsis and MSBE. These documents were then classified using our ESN model generated data network.

The classifier provided an output with the domain inclination percentage. The same documents were then manually classified by the researchers. Our system pointed out some hidden information of the newly classified documents. For instance, certain documents belonging to the original domain Arabidopsis showed inclination to other domains like MSBE a detail not specified until and unless the document is completely read by the user. This information was missed by the manually classified result. The correctness of our classifier results were verified by the domain experts who

manually verified the documents and confirmed its inclination to both the projects.

5. KNOWLEDGE MAPS USING MOLAGE TOOL- ITS APPLICATIONS ON ToxNuc-E

As the size of digital libraries (more specifically in our case; research documents shared on the platform) grows in size/ number, the lack of adapted tools for browsing in these libraries appears more and more crucial. For instance, in the context of the ToxNuc-E project, such a tool should be able to retrieve from the library not only the documents the user expects but also documents as suggested by the tool, unexpected for the user like information coming from other disciplines involved in the project, but dealing with the same scientific concepts.

Such an information retrieval tool is clearly semantic-oriented and the one we have proposed is based on knowledge maps. Building a knowledge map consists in representing contextual information, i.e. knowledge, using a visual metaphor chosen according to the future usage of the map. Here we want to represent scientific documents in the context of the nuclear toxicology domain. Hence, ontology of this domain will be represented on the map, linked to the documents.

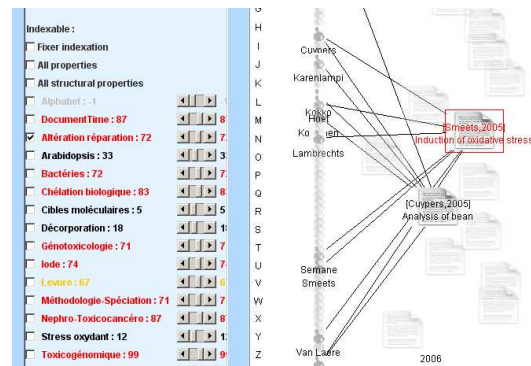


Figure 6. Display of a Navigation Path by Emphasizing Documents and Shared Concepts

Therefore, the aim is to maintain a unique mental map for the user by keeping the same visual representation of the library whenever he changes the way he browses it, rather than showing him long query result lists without any global view of the whole library. Such a representation allows the user to navigate visually into the collection of documents by different entry points (authors, ontology concepts, etc), new documents being emphasized and suggested by the tool each time the user selects a document (documents with same co-authors, linked to similar ontology concepts, ...).

institutional documents, joint workspaces, forums of exchanges, and diffusion of information to general public. Advanced functions of dynamic cartography of the contents (evolutionary trees and matrices) make it possible to follow the evolution of the data bases.

Each researcher registered in the program is a contributor authorized to deposit documents, to consult the filed documents, to communicate with the other researchers. A follow-up and a report at all stages are done every six months. We have accumulated feedbacks from the platform users on the tools provided by us and have analyzed its influence on researchers to co-ordinate and work collaboratively.

The scientific assessment of the period 2001-2003 is as follows: 79 publications with a factor of average impact of 4,17 including 6 publications with a factor of impact higher than 10 (average: 14,70); 4 articles of synthesis in referred works; 8 cards of synthesis summarizing the principal results by chemical element studied and 4 patents deposited. The human assessment is more difficult to quantify however in bringing together every Net between biologists and chemists, in the broad sense, has been achieved.

This one led for example to the appropriation of the analytical step by the biologists and of taking into account of the complexity of the world of living organisms by the chemists. Without the detailed attention of all the scientific components, technical and administrative support of the CEA, this program could not have become a success. On this same set of themes, and with the same rigor in the selection of the projects and their follow-up, we will continue the adventure with the project. The program "Environmental Nuclear Toxicology" was initiated in 2004 for three years duration by selecting the best teams of the four organizations of research partners.

REFERENCES

- [1] Penalva, J., M. Commandre "Typology of collaborative working variations around collectives in action" in Intelligence Collective, Les Presses Mines Paris, France.
- [2] Thomas, W., and K. Crowston "The Interdisciplinary Study of Coordination" ACM Computing Surveys, 1994 (March), 26 (1), 87-119, USA.
- [3] Ménager, M. "Environmental Nuclear Toxicology program: how to federate a scientific community around a societal stake" Intelligence Collective Partage et Redistribution des Savoirs, 2004 Nîmes, France.
- [4] Anoir, I. and J. Penalva "Part of confidence in collaborative collectives" Rencontres scientifiques DISTIC, 2005, Nice, France.
- [5] Hitzl, S., M. Turoff and R. Johnson "Experimentation group communications via computer, in Face to Face vs Computer conference" Human Communication Research, 13,2,P. 225-252, Human-Computer Interaction in the new Millennium by John Carrol, ACM Press, 2002.
- [6] Shetty, R. et al, "Hybrid method for knowledge processing, integration and representation" IEEE-IRI proceedings, 2006, Hawaii, USA.
- [7] Sowa, J. "Knowledge Representation: Logical, Philosophical, and Computational Foundations" Brooks Cole Publishing Co., Pacific Grove, 2000, CA, USA.
- [8] Natalya, F. and D. McGuinness "Ontology Development 101: A Guide to Creating Your First Ontology" Ontology Tutorial, Stanford University, Stanford, CA.
- [9] Sowa, J. "Conceptual structures: information processing in mind and machine" Addison-Wesley Longman Publishing Co., Inc, 1984, Boston, MA, USA.
- [10] Quillian, R. "Semantic memory" in Ed M. Minsky Semantic Information Processing, pp.216-270. Cambridge, Massachusetts: MIT Press, 1968, USA.
- [11] Brachman, J., L. McGuinness and F. Patel-Schneider "A Resnick Living with CLASSIC: When and How to Use a KL-ONE-Like Language" Special issue on implemented knowledge representation and reasoning systems Pages: 108 – 113, 1991, ACM Press, NY, USA.
- [12] Shetty, R. et al, "Hybrid Knowledge Model for Relevant Information Retrieval" KRAQ-IJCAI, 2007, Hyderabad, India.
- [13] Crampes, M. et al, "Concept Maps for Designing Adaptive Knowledge Maps, In Concept Maps" A Special Issue of Information Visualization, Volume 5, Issue 3, Guest Editors: S.-O.Tergan, T. Keller and R. Burkhard, Palgrave - Macmillan, 2006.
- [14] Ranwez, S. et al, "Ontological ISA-Distance Measure for Information Visualisation on Conceptual Maps" In On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Lecture Notes in Computer Science, publisher: Springer Berlin / Heidelberg, Volume 4278/2006, ISBN 978-3-540-48273-4, pp.1050-1061, 2006.