Extended Semantic Network for Knowledge Sharing

Reena Shetty, Pierre-Michel Riccio, Joël Quinqueton

Extended Semantic Network for Knowledge Sharing

Reena T. N. SHETTY
Doctorat, EMA, Paris
Reena.Shetty@ema.fr
Tel: +33 4 66 38 70 39
Mob: +33 6 27 42 68 85

Pierre-Michel RICCIO
Adjoint Director- LGI2P, Nimes
Pierre-Michel.Riccio@ema.fr
Tel: +33 4 66 38 70 48

Joël Quinqueton
Professor-LIRMM, Montpellier
jq@lirmm.fr
Tel: +33 4 67 41 85 85

Abstract

For ages humanity has concerned itself in fiction and often in engineering with the invention of tools meant to mime human behaviour, or to behave in an apparently intelligent way by emulating human ability to reason symbolically, as exemplified in typical Artificial Intelligence (AI) domains. It is a well established truth that web has burgeoned into a vast storehouse of various information of different forms. Since quite a time one has encountered issues questioning the efficiency of the existing search engines and the methods employed. There has been several debates on the requirement of evolution in web search techniques that are currently followed. Our research goal is to provide an efficient search tool powered by the Artificial Intelligence algorithms capable of answering the above requisitions by ably searching information and providing high precision results.

The proposition Extended Semantic Network is one such innovative tool, which not only infers meanings but looks for sets of associations between words as opposed to the present method of tag and keyword search. The objective here is to achieve a technical co-operation between mathematical models and mind models using Artificial Intelligence, thus harvesting their collective intelligence. The Artificial Intelligence tool thus developed will be capable of better understanding human mind and be able to acquire the ability of human reasoning.

Keywords: Artificial Intelligence, Extended Semantic Network, Collective Intelligence, Proximal Network, Semantic Network, User Modelling, Knowledge Management, Knowledge Sharing.

1. Introduction

Mankind has long been curious about how the mind works and fascinated by intelligent machines. One can see people's desire to understand and even to create intelligence. With today's ever accelerating advances in science and technology, it is becoming increasingly achievable that we may soon gain a complete understanding of human intelligence and consciousness. Intelligence can be described as the computational part [23] of the ability to achieve goals in the surrounding world. Varying levels and types of intelligence exist in all people, many animals and few machines.

Artificial intelligence is one such concept, which defines the science and engineering of making intelligent machines, especially intelligent computer programs capable of understanding and imitating human intelligence. AI is the area of computer science focusing on creating machines that can engage on behaviours that humans consider intelligent [12]. With this understanding it seems reasonable to assume that it will then be possible to build artificial machines whose intelligence matches, and possibly even exceeds, that of humans [2]. One can notice that the study of artificial intelligence has provided better programming techniques for building smarter computer systems.

The ability to create intelligent machines has intrigued humans since ancient times and today with the advent of the computer and 50 years of research into AI programming techniques [6], the dream of smart machines is becoming a reality. Researchers are creating systems which can mimic human thought, understand speech, beat the best human chess player, and countless other feats never before possible.

For any organisation, the single most important asset must be its individual members and their accumulated knowledge. It has been suggested that the sum of individual parts of knowledge may be less than the whole, and so methods of accessing, utilising, sharing and storing this knowledge are important factors to be addressed in any organisational studies.

Collaborative work and sharing is relatively a developing area which has introduced a methodology for the planned capture and re-use of organisational knowledge [1]. Successful application of collaboration involves the understanding and constructive use of organisational learning and information flows within the organisation. Collaborative work is becoming more important in the evolving context of global network, which places the user at the centre of a collective device [4].

In order to overcome the problem of finding, extracting, managing or sharing relevant information out of the enormous amount of data collectively available, various technologies have been explored with in the AI community.

This paper presents one such collaborative AI tool which facilitate knowledge sharing and knowledge retrieval among a vast community of researchers involved in various collaborative fields of research activities. Alternatively, our research aim is to find a new and optimal solution to achieve collective intelligence based on the attributes of artificial intelligence tools.

Here, we present a new method of combining the concepts of semantic networks and proximal networks to be used in knowledge management, extraction and sharing of relevant information. We investigate the possibility of combining the different methods of knowledge extraction and knowledge sharing to form one such tool called *Extended Semantic Network*. Thus, leading to an effort to develop a tool harvesting the collect intelligence [4] of these different methods using the AI techniques.

The following sections are organised as follows: In section 2 we review related work in the field of information sharing, search and retrieval using artificial intelligence. In section 3, we propose novel approaches for boosting collaborative working environment by encouraging knowledge sharing and easy retrieval, followed by the results of our approach, advantages and future work. Finally, we conclude

this paper in section 4 with the summary of the work completed and direction for future work.

2. Related work

Previous research and recent developments have highlighted the importance of collective intelligence and its advantages. It wasn't so long ago that the waggle dance of the honey bee, the nest-building of the social wasp, and the construction of the termite mound were considered as somewhat magical aspects of nature.

How could these seemingly uncommunicative, certainly very simple creatures be responsible for such high feats of organisation and creativity remained a mystery [3] Over the last fifty years biologists have unravelled many of the mysteries surrounding social insects, and the last decade has seen an explosion of research in fields variously referred to as collective intelligence [19].

Here, we attempt in combining the features of human reasoning and mathematical topology to create an innovative artefact which would facilitate information sharing and search.

2.1. Web Search and The Existing Approaches

The web searching phenomenon started by the mid nineties has become cumbersome. Web has burgeoned into a vast storehouse of information, with hundreds of pages added each day on thousands of topics. Yet for someone seeking reliable information on a particular topic, the library is still a better option, due to the lack of effective search tools. Internet search engines of the time relies completely on techniques from information retrieval systems to find documents that contains the search terms prominently and with high frequency.

But these tools are primarily designed for databases with limited size, consisting of high-quality information presented in a controlled style and structure [25].

2.1.1. Page Ranking

The Web, with its base structure, is a chaotic jumble of pages of variable quality offered in a variety of formats. The IR techniques alone would rank junk pages highly if they happened to contain multiple instances of the search term. At the same time, the techniques would miss the relevant pages that did not include the term at all. A search for “Cars” or “automobile manufacturers” would not return the website of companies like Honda or Ford.

2.1.2. Keyword Search

Simple keyword searches has another disadvantage: people wishing to draw more traffic to their web sites could easily exploit them in their gain. The only hope for extracting useful information from the web is to have an army of human researchers [13] play the role of a librarian, shifting through the clutter and selecting and organising the pages based on each subject or category. But with the web adding thousands of pages a day, such directories are necessarily limited in scope.

One of the approaches, offered by two Stanford graduate students became the basis for goggle. The other approach was the one developed by researchers at IBM, which has been extensively used in customised search applications. Both methods are highly mathematical, requiring real-time computations on enormous data sets that grow with size of the web. Applied mathematicians are now rising to the challenge, finding more efficient ways to implement these complex ranking algorithms.

2.1.3. The Semantic Shoe Search Engine

The basic idea of this search engine is that all pages indexed by the search engine must use a special set of tags, sort of meta-tags on steroids. These tags provide more than keywords and descriptions, they describe content and relationships. Like in XML, more and more web pages are created using content-oriented tags as opposed to the presentation-oriented tags provided by HTML.

2.2. Methods Compared:

For significant fraction of queries the above methods differ in interesting ways. The methods are computationally distinct as well. The Brin-Page approach uses a single ranking of all Web pages to return appropriate results for a search, while the Kleinberg method customizes its ranking to each search. While Page Rank is query-independent, google can compute offline, updating only once every month or so to reflect the growth and evolution of the Web.

Kleinberg's method requires a real-time ranking computation following each search. At first this seemed to be a prohibitive barrier to its use in a large scale search engine. Another downside to Kleinberg's methods is its vulnerability to spamming. A person can influence the authorities score of a site just by adding links to popular sites. Having established a few good authoritative they can then use them to improve the authority score of any site they choose.

However, the shoe solution works by associating a context with a Web page, this context can be used to disambiguate terms and provide background knowledge that might help in interpreting content. The shoe calls this context an ontology, which is just a fancy way to say a vocabulary and what it means.

But suppose they require a search engine that would look for a set of associations not just a literal tag, the *Extended Semantic Network* sort of does this by linking associated to one another. Here, it links all the associated words to the one the user is looking for.

3. Our Proposal--Extended Semantic Network (ESN)

3.1. General View

Extended Semantic Network as the name suggests is a result of collaboration between automatically constructed proximal network and manually constructed semantic network. Here, the primary idea is to develop an algorithm combining the features of different networks, which can be of enormous use in the latest computing, pattern matching and collaborative tools. We propose to visualise a novel method where the representation is partly based on mind modelling and partly on the topological method.

In ESN we endeavour to construct a network based on human constructed network, utilising the statistical data obtained by our mathematical models. This in turn generates a tool which can help in optimising, information search and subject classification procedures, enabling group interaction in a society. The ESN forms a hybrid by inheriting the features of both the source networks computed differently and independently,

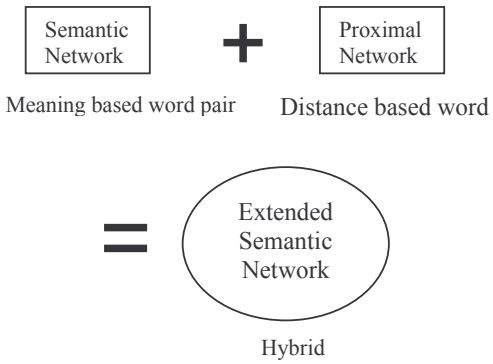

Figure1. Schematic Representation of ENS

making it a more robust and optimal algorithm.

3.1.1. Semantic Network Prototype

Semantic network [8] is a labelled, directed graph with nodes representing physical or conceptual objects and labelled arcs representing relations between objects. This permits the use of generic rules, inheritance, and object-oriented programming. A semantic network is often used as a form of knowledge representation. It is a directed graph consisting of vertices, which represent concepts and edges, representing semantic relations between the concepts.

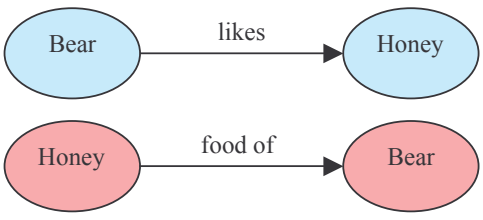

Figure 2: Single-labelled Semantic Relation

Semantic networks are a knowledge representation technique. More specifically, it is a way of recording all the relevant relationships between members of a set of objects and types. By "Object" we mean an individual and by "Type" we mean a set of related objects [9].

Here is a pair of labelled nodes and a single labelled edge relationship between them. More over there could be more than one relationship between a single pair of connected words: for instance the relationship is not symmetrical and there can also exist relationship between the above nodes through other indirect paths [6].

Below is a larger fragment of a semantic net, showing 4 labelled nodes and three labelled edges between them.

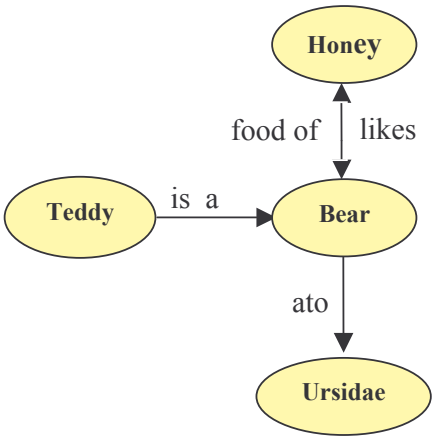

Figure 3 : Multi-labelled Semantic Relation

Technically a semantic network is a node- and edge-labelled directed graph, and it is frequently depicted that way. The scope of the semantic network is broad, allowing for the semantic categorization of a wide range of terminology in multiple domains.

Major groupings of semantic types include organisms, anatomical structures, biologic function, chemicals, events, physical objects, and concepts or ideas. The links between the semantic types provide the structure for the network and represent important relationships.

In our semantic network prototype all word relations are built based on the sense or the meaning that the word pair share with a possibility of more than one relationship between a single pair of connected words. All the links used in connecting the word pair here is based on the UML links consisting of four different types of associative lines. The semantic network can change depending on the individual opinion and understanding.

The four types of UML links used by us are as shown below. They have been chosen after several experimentations, based on the requirements of our algorithm.

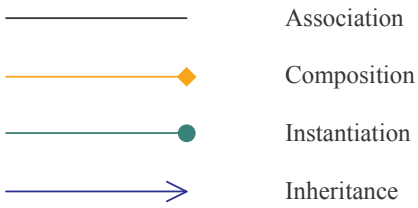

Figure 4: Different kinds of Links used for Constructing the Semantic Network

3.1.2. Proximal Network Prototype

Proximity is the ability of a person or 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 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. If entities are logically positioned they connect to form a structural hierarchy.

The proximal network Prototype model is built based on the proximity of words, documents and the information to be shared [13]. Here proximity is calculated only considering the physical distance between the occurrences of the above entities at a given instance. It is the property of being close together considering the distance shared between them. The proximal network prototype developed is completely based

on the statistical calculations using different algorithms. Here we use UML link of association to connect words or nodes proximally closer.

Results obtained from the semantic network are considered as the core of our network based on which the ESN network will be constructed. We extend the results of semantic network by adding on the results obtained by the proximal network thus making it an *Extended Semantic Network*. The demonstrable prototype of ESN has been developed based on the data of ToxNuc-E project [14].

3.2. Application on Environmental Nuclear Toxicology

In *Extended Semantic Network* each of the sub-networks are constructed separately using different methods and are then combined to obtain the ESN network with a better performance ratio, regarding a given criterion (to be described) as compared to its sub-networks.

The documents used in our experimental prototype are chosen from a research website called ToxNuc-E [14], a website devoted to all the research activities carried out in Biological, Chemical and Nuclear domains. It is a platform where researchers from different domains, meet and exchange their views on various on-going research activities. The ToxNuc-E currently has around 660 researchers registered with their profile, background and area of their research activity.

3.2.1.Semantic Network

Our Semantic prototype is being developed based on the opinion of a set of specialist in each of the chosen domains. To start out with, we have selected a set of 50 concepts concerning to each of the chosen subject or domain. We then consulted people who were either specialists or who possessed a high level of knowledge in chosen area of study.

These people were provided with the concept list on which they were requested to develop a network depending on their individual view point. The word network developed by each of them was then analysed and merged to obtain a semantic network for that domain.

This process was repeated on different lists of words related to various domains. The network was then constructed by linking the word pair using four different types of links as explained earlier.

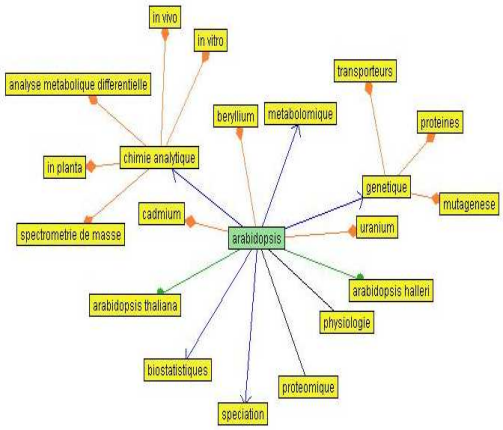

Figure 5: Semantic representation of Concepts using Graph editor

The semantic network thus generated is then stored into the MySql database and are utilised using a graph editor developed by us using java application.

3.2.2. Proximal Network

This network is constructed using documents pertaining to the numerous research activities being carried out in the field of nuclear toxicology in plants and animals. These documents are pre-treated to obtain a matrix of words and documents as row and columns respectively. Here, java has been used as the programming language and all the data in use are stored using the MySql database.

In this network, we are primarily concerned with the physical distance that separates words and documents in a given space. Currently, we have been able to successfully process around 3423 words and calculate the physical distance between them using various mathematical algorithms and calculations. This prototype is also

bilingual with the ability to distinguish French and English words before processing them.

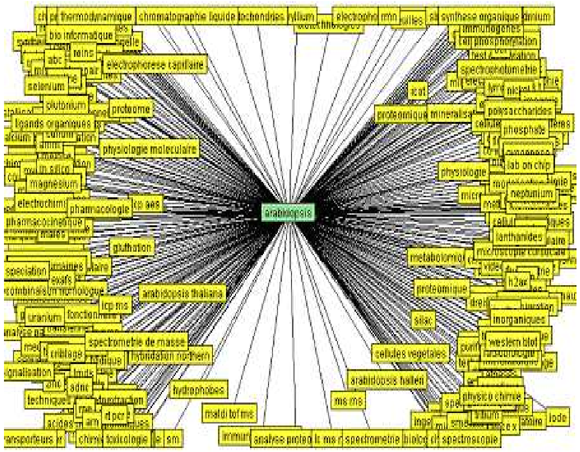

Figure 6 : Graphical Representation of Proximal Network using Graph Editor

The result obtained from the proximal network is then fed into a Graph Editor for graphical visualisation. The graph editor displays all the results by displaying a value between every word pair when required.

Our approach relies on numerous ideas and details which have been established in previous research works by various researchers. The results are stored into the database to be used in combining it with the results of the semantic network.

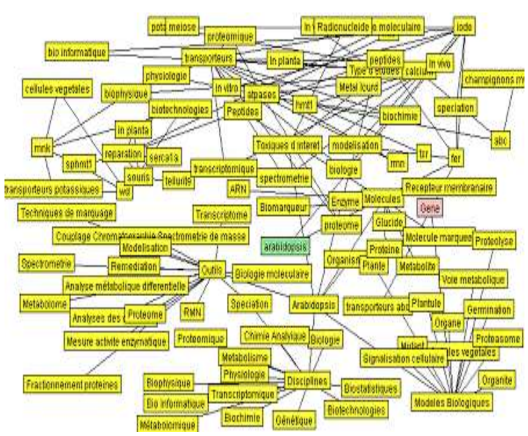

Figure 7: Extended Semantic representation of Concepts using Graph editor

Currently, the 2 different results are combined with simple extension methods. Simultaneously, several other optimising algorithms are being considered to be utilised in merging the networks to build the *Extended Semantic Network*. We are exploring the possibilities of using the genetic algorithms and features of neural networks to obtain an optimal result. Our results have been verified by experts and validated for providing satisfactory results. The ESN thus targets at initialising a new generation search algorithm to facilitate information search, retrieval and sharing.

This makes it efficient and a very reliable method for calculating the word/document association and then sharing them over an electronic media.

3.3. Advantages and Future Work

The results of proximal network and semantic network have been tested separately with several data and have been verified by experts. We are on the conclusion that the results are exceedingly reliable and accurate. Our next step will be to experiment different algorithms to converge these results into one network the *Extended Semantic Network*. The aim is to develop an application library based on the results of *Extended Semantic Network*. This application library is intended to be used independently for various applications like web search, document classifier and/or as a tool for information exchange on a collaborative platform. This will be used for data and knowledge sharing and retrieval by the researches working with Toxnuc-e.

The advantages of *Extended Semantic Network* are of enormous importance in the field of artificial intelligence in terms of mind model research. We will carry on our research in the field to produce a collaborative knowledge sharing and retrieval mechanism which can think in similar grounds as of human mind and also be as accurate as possible in comparison to the later.

Once we achieve this, we will start on our experimental research in user modelling. Below is a small introduction to user modelling

3.3.1.User Modelling

User Modelling has been found to enhance the effectiveness and/or usability of software systems in a wide variety of situations. A user model [15] is an explicit representation of properties of a particular user.

A system that constructs and consults user models can adapt diverse aspects of its performance to individual users. The Techniques for user modelling have been developed and evaluated by researchers in a number of fields, including artificial intelligence, education, psychology, linguistics, human-computer interaction, and information science.

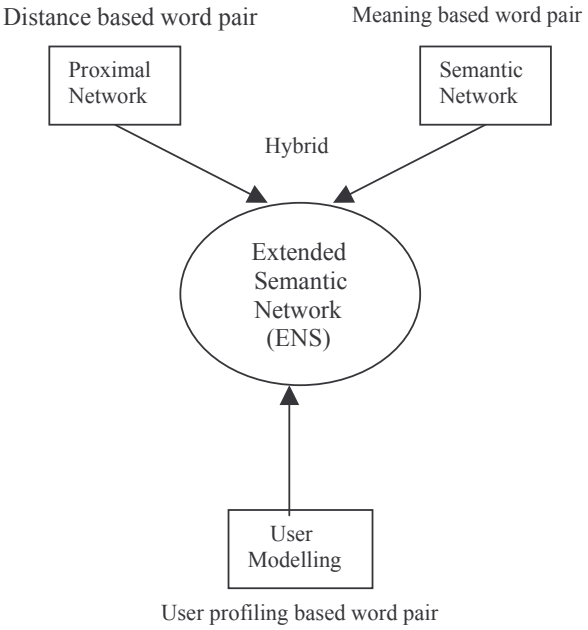

Figure3. Schematic Representation of ENS

Information about a user can allow a human assistant to provide high quality support by tailoring the consultation to the individual user. Having such information readily available can also make consultation dialogues more efficient and thus save time and other resources. Here, the system adapts to the requirements of each user based on the user profile. We intend to upgrade our first prototype of *Extended Semantic Network* with the attributes of user modelling thus customising are algorithm for each user profile.

User profile for each user is created and stored in a database based on the user history. This database is updated on a regular basis with the contents that the user in question is currently interested. Similarly, we plan to monitor the behaviour; 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 management, sharing and retrieval. There is yet a lot of research to be done in this domain by us before cornering the optimal solution for finding the most relevant results.

4. CONCLUSION

For centuries various developments in the world of artificial intelligence have fascinated and challenged the intelligence posed by us humans. There have been various evolutions in machine and computers around us. Current advancements have taken humans to places and heights that were never believed can be achievable by humans.

The amazing capability of people to learn and recognize things has been taken for granted for years. It is until recently, when one tries to use computers or machines to do things like recognizing handwritten characters or faces, it becomes clear that such seemingly trivial tasks by human being turns out extremely difficult, if not impossible, by mechanical means such as computers.

Even though the techniques of artificial intelligence dates back to centuries old, real term advancement was carried out during the mid 19th century. Scientists around the world began to identify various stages of success one would achieve with the AI techniques. These researchers led to various inventions in the world of AI techniques. Some of these researches have been a great boon to the human race while some have been cause to destructions.

The question on knowledge management, sharing and data retrieval is both fascinating and

complex, essentially with the co-emergence between man and machine. This research paper presents a novel collaborative working method, specifically in the context of knowledge management, retrieval and sharing. The advantages of our methodology with respect to the previous work, is our innovative approach of combining machine calculations with human reasoning abilities [24].

We use the precise, non estimated results provided by human expertise in case of semantic network and then merge it with the machine calculated knowledge from proximal results. The fact that we try to combine results from two different aspects forms one of the most interesting features of our current research We view our result as structured by mind [23] and calculated by machines. Our future work would be to identify the right combination between the two vast methods and setting up a benchmark to measure our prototype efficiency.

But with the ever growing Human Intelligence, there is no room for attaining saturation in the field of artificial intelligence. This explicitly implies the need for a more robust and a dedicated research in artificial intelligence to carry it on to its next level or generation.

5. REFERENCES

[1] M Commandré, J.M Penalva Typologie du travail collaboratif : variations autour des collectifs en action, Intelligence Collective Partage et Redistribution des Savoirs, Nimes, France, septembre, 2004.

[2] J Quinqueton, Aspects socio-organisationnels dans les systémes multi-agents : l'intelligence artificielle en essaim, Intelligence Collective Partage et Redistribution des Savoirs, Nimes, France, septembre, 2004.

[3] J Ferber, Socio-organizational aspects in multi-agent systems : a generic model of organisations for structuring MAS, Intelligence Collective Partage et Redistribution des Savoirs, Nimes, France, septembre, 2004.

[4] S Iksal et S Garlatti, Adaptive web information systems: Architecture and Methodology for Reusing Content, Intelligence Collective Partage et Redistribution des Savoirs, Nimes, France, septembre, 2004.

[5] L.M Garshol , What Are Topic Maps?, ISO/IEC JTC1/SC34 300–350, September 11, 2002.

[6] J.F Sowa , Knowledge Representation: Logical, Philosophical, and Computational Foundations, Brooks Cole Publishing Co., Pacific Grove, CA, 2000.

[7] J Voss, P Danowski, B Chapter - citebase.eprints.org, 2005.

[8] M.R Quillian,, Semantic memory. M Minsky, Ed. Semantic Information Processing. pp.216-270. Cambridge, Massachusetts: MIT Press, 1968.

[9] J.F Sowa, Conceptual structures: information processing in mind and machine, Addison-Wesley Longman Publishing Co., Inc, Boston, MA, 1984.

[10] J Brachman, L Deborah, McGuinness, F Patel-Schneider, A Resnick Living with CLASSIC: When and How to Use a KL-ONE-Like Language, 1991.

[11] Rational Corporation: UML Notation Guide 2, 2000.

[12] M.E Winston, R Chaffin and D Hernnann, A taxonomy of part – Whole Relations Cognitive Science 11, 1987.

[13] S.A. Mahé, P.M. Riccio et S. Vailliès: des elements pour un modèle: la lutte des classes! Revue Génie Logiciel, n°58, Paris, septembre 2001.

[14] M Ménager, Programme Toxicologie Nucléaire Environnementale : Comment fédérer et créer une communauté scientifique autour d'un enjue de société , Intelligence Collective Partage et Redistribution des Savoirs, Nimes, France, septembre, 2004.

[15] J Aberg & N Shahmehri, User Modelling an Aid for Human Web Assistants, User Modeling 2001: 8th International Conference, UM 2001, Southaven, Germany, July 13-17, 2001.

[16] E Reingold, J Nightingale, "Artificial Intelligence".

[17] Alexander maedche & Steffen Staab, "Ontology Learning for the Semantic Web".

[18] J Link-Pezet., P Glize, C Régis,, A cognitive approach to intelligent databases. On line, London: Learned Information, 1992.

[19] N.J Belkin, W.B Croft, Information Filtering and Information Retrieval: Two Sides of the Same Coin?, Communications of the ACM Vol. 35 n°12, 1992
[20] R Davis, B.G Buchanann, Meta-Level knowledge: Overview and applications, IJCAI, ACM SIGIR, n° 5, Cambridge, 1984.
[21] A Maedche, S Staab, Representation & Learning, IEEE Intelligent Systems, 2001.
[22] E Rosch and B. Mervis Family Resemblances: Studies in the Internal Structure of Categories , University of California, Berkeley, 1989
[23] E Rosch Cognitive Representation of Semantic Categories, University of California, Berkeley,1978
[24] E Rosch " Cognitive Reference Points", University of California, Berkeley, 1978
[25] E Rosch, B. Mervis, D. Gray and P Boyes-Braem Basic Objects in Natural Categories, Cogn. Psychol., 8:382--439, 1976.