Chaos, Cognition and Cyberculture – an environment

Marcos Luiz Mucheroni[1], Ro Yih Ching [2] and Cairo Gomide Jr [3]

Abstract — *The representation´s concepts of knowledge and semantic databases can be applied to object orientated concepts to knowledge, such as, objects and classes of information in properties representation and relationships of them. This work is an environment project to provide applications of using knowledge-based data, organized in Content Management and user interface in an self-organizing mode. It includes considerations about knowledge, networking and formats. There are brief theoretical considerations about knowledge and intelligence, to justify the need of this chaotic system, described soon afterwards. The data is put in matrix model and used the algorithm SOM (Self Organizing Maps) in a syntax for the representation called RDF (Resource Description Format - an extension of MCF) in XML (EXtensible Markup Language) and a vocabulary (or descriptors group) for the Internet. The use of the language Java will give support for the metadatas models and interoperability among several patterns. Measurements with 100 virtual users were made.*

Index Terms — Chaotic database system, knowledge-based, cognition.

1 - INTRODUCTION

Most of the theories and humans believes (the believes are nearer to us than what is imagined) are codified by a cultural compound creating a semantic set that includes its language and its knowledge, as well as the language is related to the code, knowledge is related to information but the intelligence is something superior capable to releasing men from these conditioning and build something really new, and this is what is called here current intelligence, in narrow connection with Pierre Lévy´s [14] concepts of actual (but not the real) opposing the virtual.

Language, codes and codification can be considered as fundamental part of what called reality, but this reality is different from the abstract reality that it is more related to the knowledge. Of course, we have an entire philosophical discussion here on reality and knowledge. In the practical sense in this work, knowledge is used as Levy´s definitions: "Knowledge, values and tools transmitted by the culture..." , "But always...of a dialogue or of a multi-dialogue, real or imaginary" and ""we will end up in the instruments... it is impossible to exercise our intelligence independently of language, vocabulary and systems of signs" [13].

In this subject, we speak about certain levels of abstractions and of realities, or more concretely as defines by Pierre Lévy [13], intelligence levels linked at certain levels of technologies, but like he says technology is "classification to help to locate the poles", what can be made not only in relation to technology, but also to the productive development, the complexity of the language (orality levels, for Lévy only two) and why not, the level of elaboration that get together man and nature as a complex life, that it is our thought.

Apprenticeships different from the man allowed different orality forms, code, languages and thought styles, and for its time they generated differentiated forms of storage and classification of the information. For the current cognitive sciences [17] they are called of code strategies [1] and but they also present the result that the conscience no resulted of the language [3], and many errors and fallacies about consciousness and intentionality [5].

But this doesn't mean that language and code are not important, and it is possible to maximize its potential in the information age, we will need a better understanding of nature intelligence and technological knowledge in matter, and the human concepts and its limits in general.

Recent researches give some hints that certain mechanisms of the brain work in a self-organizing way [10].

The elaboration in this first part of the work is to build a digital system of easy access and of great interconnection capacity for networking, and for that is necessary to choose a storage format, proposed with the name Small Embedded of Data Center Environment "Cognitio" (SEDEC).

The system uses feedback information from its users to change its representation about the data (e.g., in digital library: about of authors, index terms, and documents), is possible the users to change its representation of data, index terms, and documents over time, so this include cognition on the system, so "Cognitio" is the name. Many systems are developed using a similar model of three-layer network of queries [12] also, index terms, and documents and others use modified Hebbian learning rule [7] which was used to reformulate probabilistic information retrieval [19].

While the above systems represent information retrieval applications in terms of their main components of documents, queries, index terms, authors, etc. other researchers have used different neural networks for more specific tasks. The "embedded" is because any applications will be included in target system (e.g. digital library).

[1] Marcos Luiz Mucheroni, Euripides Foundation of Marília, Av. Hygino Muzzi Filho, 529 – 17.525-901, Marília, SP, Brazil, mucheron@fundanet.br
[2] Ro Yih Ching, FIM student Euripides Foundation of Marília, Av. Hygino Muzzi Filho, 529 – 17.525-901, Marília, SP, Brazil, ching@ieg.com.br
[3] Cairo Gomide Jr, student Euripides Foundation of Marília, Av. Hygino Muzzi Filho, 529 – 17.525-901, Marília, SP, Brazil, cairo@fundanet.br

Mind, cognition and consciousness

Good part of the contemporary knowledge leaves concrete for the abstract, as if the theory was the arrival point. But the new scientific paradigms make a more complex way, they leave the abstractions to elaborate other abstractions, worst, because we always have multiple theoretical points of departures and they turn the theory arrival points, while the reality, even if not completely understood (the real intelligence) it is always closer of the human problem than the abstraction.

It is right that the knowledge comes from reality of the real (not actual), but the real has different reality levels, for simple example, four apples are 4, but what is 4? - an abstraction and a sign! Then remove the quantitative factor: an apple will be a code to indicate a group of complex qualities of certain fruit, in agreement with the intelligence of who hears he translates it in a certain representation level, or abstraction level.

Cognition is defined in the dictionaries as: "the action or process of knowing in the widest sense; specifically, an intellectual process for which there is a knowledge earnings, of perception or ideas" and in a certain way is quite compatible with the Hebbian neural learning systems [7].

But, is the mind that translates the world of particles and waves in a world of colors, sounds and smells, how? If we know the answer that can make the machines imitate our human mind's processes and we can put them to aid our work.

The psychology, philosophy, information's theories, the cognitive sciences and the neurophysiology tried this to answer while the artificial intelligence initially tried its own road, but today its is closer to this because of the Internet and the development of a cyberculture.

It is possible to simplify move the discussion foreward, defining mind as:

Mind = cognition + consciousness

where: the cognition is omnipresent in the nature, and consciousness is a natural phenomenon, present in the man as subjectivity.

But the remaining question is: "How do brains work?" The studies of technical tools [18] and others enables us to model brain functions that transcend the present limitations of computational or schematic models, and go to the domains of non-cartesian concepts of consciousness emerges, any insight was arrived in past century: the cycle of "action, assimilation and adaptation" [18], "Mind is the structure of behavior" [8] or the ecopsycology of the "an affordance ... of an object offers what it does because it is what it is" [6].

Knowledge in technological information age

As well as the primary orality and the writing they generate a relationship different with the knowledge, the cyberspace supports intellectual technologies that enlarge, they utter and they alter many human cognitive functions: the memory (databases, hipertexts, digital files [numeric] of many types and orders), the imagination (simulations), the perception (sensor digital, telepresence, virtual realities), the reasonings (artificial intelligence, neural networks or other model of complex phenomena).

The actual cyberspace is Web, but the Web is still very rigid, it offers a group of options but the user's possibility to create their own ones and stiller, of to show his intention and to ask the machine that helps him in the development of his insight is still a distant possibility.

The elaboration in this first part of the work is to build a digital system of easy access and of great interconnection capacity for networking, and for that is necessary to choose a storage format.

There is a large number of different formats of different kinds of information can be stored, but knowledge includes more general information such as: voice, images, songs, feelings, or simply a conversation with a friend to be registered as a note or release. Therefore the best format doesn't exist nor one best piece of software to access the data in general the document is served as different information source about a number of different formats. Like this we don't intend to define an unique format of input, but a concise formatting that gives the maximum flexibility to the user that makes the consultation as a possible source. It is intended besides to allow, at the interface level, that the user organizes his search in a personal way, and the group of this search will be in chaotic mode using self-organization.

Chaos and self-organizing information

The chaotic mode is not disarranged or disturbed organization, or its organization named self-organizing, it is a complex way of organization and it is compatible with a new science: Chaos Theory.

This theory is specially applied to non-linear system in different techniques, one technique is to detect special singularities named "atractors" in time series analysis.

In this work the concept of special method of self-organization named SOM (Self-Organizing Maps) is use to search words and texts, tags and special words in Web environment and special data in Database Content. Metadada in the environment compounds this Data Content.

Lin [15] adopted a Kohonen network for information retrieval. Kohonen's [11] self-organizing feature map (SOM), which produced a two-dimensional grid representation for many features, was applied to construct a self-organizing (unsupervised learning), visual representation of the semantic relationships between input documents. Many type of SOM algorithms is being built, but a special algorithm was made for to searching in the WEB, and is named WEBSOM [9].

2- AN ENVIRONMENT – THE SEDEC

Many companies, organizations and several enterprises that use digital information that use on-line digital systems

and are now upgrading their systems to WEB systems [16] or beginning a new Web system, e.g. e-commerce system.

But now any new characteristics are happening and new problems are detected:

- a lot of information is on-line in the net, but without patterns and a lot of duplicated "lineal" information (e..g. letters, emblems, gifs, images)
- many light devices are available (e.g. Palms, PDAs.), but the connection with information is difficult.
- there is a lot of memory available in the net, but the security and administration is critical.

The first problem is standard and the new concepts that discussed more and more this problem to search more and more patterns, but the reality sense is use any special patterns in an special box named container. This concept appear in programming Java language in connection to components, that is the element that add to a container are tracked in a list.

The second, lineal information is stored in two types of documents: Content Management to typical texts, emblems, etc. and Backup/Reporting to schedulers, securities info, the copies or other needs. This information is related to Content Database, that is, the metadata where any patterns is related and is provided to add or exclude any field, tags or labels to this metadata if the any point manages. One very fast station control this Content Database, and two others station treat small (e.g.text) and big data (e.g.images).

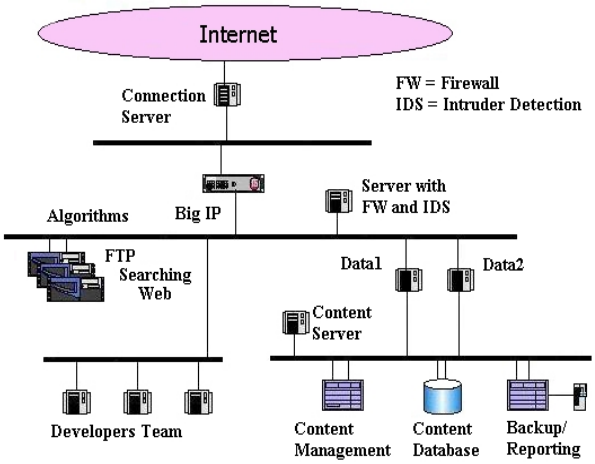

FIGURE. 1 –A SEDEC "COGNITIO" ENVIRONMENT.

The security problem is solved using Firewalls (FW) and IDS (Intruder Detection Systems) in a system named in-depth model, because track each connection traversing all the firewall's interface to make sure they are valid. As FIGURE 1 shows, is a partial system because it lacks a system of applications.

Special devices: downloads, search and WEB

The special effort for this system is the use of small Data Centers geographically separated, that is, distributed, and this result new problems: downloads, searching and WEB organization. Other problem cited later, is the WEB offers a group of rigid options and doesn't has flexibility for user search his proper paths.

The special device Big-IP is to perform local load-balancing (LLB). The machines provide load-balanced IP address to the Internet and distribute incoming request locally based on any server's availability and response times.

For primary stage in this implementation, we installed two device named Data1 and Data2 in the same place, for academic reason, but in application target we think implementation in data centers distributed and geographically separated. Three primary facts drove our goals: ease of management, low costs and democratizes the resources. The main goal is distributed data centers.

Any organization's requirements was evaluated: trace ability to a common problem, a document is commonly the final goal of activities in business and research but has many complex documents (multimedia), documents as corporate or group memory and others.

A common technique in linear system is to use linear, random or stochastic mapping for information organization to searching and sorting, included: sorting method (bubble or shell, for example), neural networking and other. But ordinary people, which are unaware of these methods, have a proper way to organize information, that realization motivates the development of self-organized systems.

Three software devices are proposed: WEB, FTP/ Downloading and Searching. Current FTP is made in software, but in target system will be in FPGA devices.

An important requirements is the management of different formats, for this we determined that the Content Management Application (CMA) should reside wholly in staging. The use of XML and the concepts of RDF (Resource Description Format - an extension of MCF) in XML (EXtensible Markup Language) and a vocabulary (or descriptors group) is done to solve these problems.

Searching WEB using the SOM algorithm and special search (e.g. using descriptors group), guarantees good results, but he main advantage is that distributed centers can have searches self-organized done in each user's owner formatting. The final implementation will provide the own interfaces. Separate users (or servers in goal application) at each location handle all available devices and backups/reporting, or documents with Content Management.

In the heart of this problems is a SOM (self-organizing algorithm), and initial evaluation is to simulate this devices environment, tests and evaluations they can prove the need that these devices are made in hardware, in this case, we will opt for algorithms in FPGAs.

The self-organizing map (SOM)

The self-organizing map (SOM), developed by Teuvo Kohonen in the decade of 80, present a map topologically based in the cerebral cortex. It is known that the brain of the most sophisticated animals has areas that are

responsible for specific functions. Areas exist, for example, dedicated to the speech, to the vision, to the motor control, to the sensibility, to the touch, etc. Each one of these areas contains subareas (each subarea map is responsible of the sensorial organ represented by the map). For example, cortex auditory map capture the different sound frequencies and cortex visual map capture visual primitive characteristic, such as light intensity, orientation and curvature of lines [2].

As all the artificial neural network, the Kohonen's map are formed by a group of simple elements, called neurons, organized in more complex structures, that work together: the net (or map). Each neuron is an unit of processing that receives incentives (of out of the system or of another neurons), and it produces an answer (for another neurons or outside of the system). Just as the neurons of the brain, the one of the neural network are interconnected itself through ramifications which the incentives are spread. Every entrance is connected to all the output neurons.

A Kohonen's model is based in a competition of the information representation presented in the input data by output layer neurons. Choose a winner neuron, this is readjusted to answer still better to the received incentive. Inside of this not supervised model, not only the winner, but also its neighbors (in a physical sense) are adjusted.

Two important aspect of data are developed in SEDEC: storage, management and organization using metadada in Data Content Management, and a second aspect, the relationship between related data for intelligent search and research using SOM algorithms.

3- THE ENVIRONMENT (SEDEC)

In first aspect (Data Content Management), is important the use of RDF (Resource Description Format) and XML.

Starting from the model MCF (Meta Content Format), RDF created is a Formal Model of Data and it can be defined like "Architecture Components to Support Metadata Interoperability", with a transfer syntax based on XML, being this a Syntax human-readable and machine-processable, or be allowed each description community (in our case the user) to define its own semantics, without the need of an only pattern for all, like this she provides a structural interoperability. This model uses graph addressed and a tool (Resource) that is described through a collection of properties (Property.) call of RDF Description, where each property has a Property Type and Value. Through the property Type the resource obtains a value. URI is a value constant defined using strings. The FIGURE 3 shows RDF:

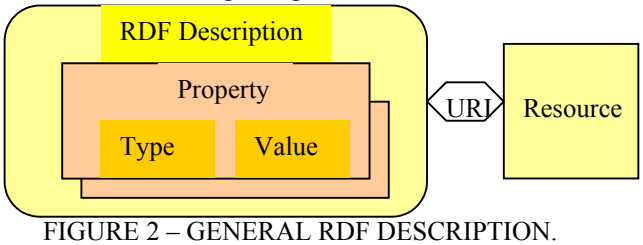

FIGURE 2 – GENERAL RDF DESCRIPTION.

This neighborhood region can assume several different formats. Although the square format is the most common, the neighborhood area can also present, as in the FIGURE 3, the form of a hexagon or circle (gaussian). The definition of the most appropriate format depends on the problem and the data's distribution.

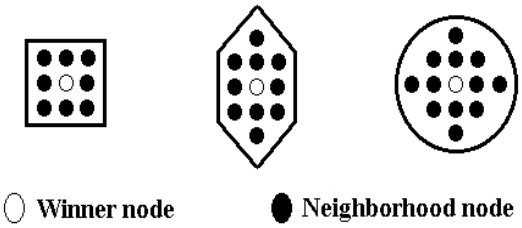

FIGURE 3–DIFFERENT FORMAT OF NEIGHBORHOOD

To understand how the algorithm works, consider the follow input data, the weights and the output neurons in the FIGURE 4. For each input data, has a weight, and each input data is linked to each output neuron, so W_{ij} defines the weight, where i represents the entry X and j the output neuron [4].

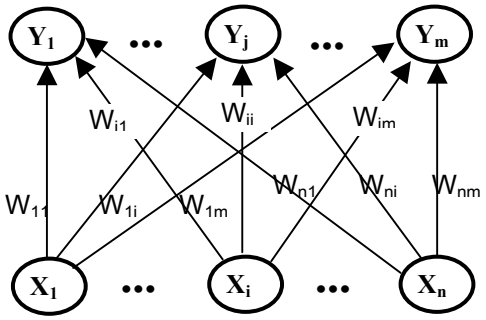

FIGURE 4 – EACH INPUTA DATA X IS CONNECTED TO OUTPUT NEURON Y.

The neighborhood region, or topology, of self-organized map, can be square, hexagon, circle, like in the FIGURE 3, for a didactic reason, a lineal topology will be used, shown in the FIGURE 5. The nodes (or neurons) they are represented for *, and the winner node for #. The rays of action of the winner node is illustrated by symbols: {}, (), [].

* * * { * (* [#] *) * } * *

{ } R = 2 ( ) R = 1 [ ] R = 0

FIGURE 5 - LINEAL TOPOLOGY .

Consider a weight Wij, an input vector Xi, a group of output nodes Yj and the learning rate α that can be modified by each group of entrance vector, so step-to-step:

0. Initialize weights Wij.
   0.1. Set topological neighborhood parameters.
   0.2. Set learning rate parameters (α).
1. While stopping condition is false, do step 2-8.
   2. For input vector x, do step 3-5.
      3. For each j, compute:

$$D(j) = \sum_i (W_{ij} - X_i)^2$$

4. Find index j such that D(j) is a minimum.
5. For all units j within a specified neighborhood of Yj, and for all i:

$$Wij(new) = Wij(old) + \alpha[Xi - Wij(old)].$$

6. Update learning rate.
7. Reduce radius of topological neighborhood at times specified.
8. Test stopping condition.

The learning rate α is a slowly decreasing function of time (or training epochs). Kohonen indicates that a linearly decreasing function is satisfactory for practical computations; a geometric decrease would produce similar results [4]. The radius of the neighborhood around a cluster unit also decreases as the clustering process progresses.

The formation of a map occurs in two phases: the initial formation of the correct order and the final convergence. The second phase takes much longer than the first and requires a small value for the learning rate. Many iterations through the training set may be necessary, at least in some applications.

4 – RESULTS AND CONCLUSION

Before going live with the initial infrastructure, we realize tests to verify that it met necessary performance requirements and to remove any possible bottlenecks.

Testing simulating up to 100 simultaneous virtual users in intranet (the most interesting is use modem but for the time being we don't have this possibility), using a set of virtual users to travel specific paths of varying lengths through the site and download documents and data of various sizes and formats.

The Table 1 illustrates, the primary resources linear and primary resources used in self-organizing model, using transfer rate of the 100 Mbps in FastEthernet, the storage progression (in Mbytes by minute) for two machines Data1 (simple texts) and Data2 (images), the storage in minutes is:

TABLE 1

Storage	1 Virt Users	5 Virt Users	50 Virt.Users	100 Virt.Users
Data1	33.4 Mb	9.45 Mb	4.77 Mb	2.12 Mb
Data2	33.2 Mb	8.40 Mb	4.20 Mb	1.86 Mb

The conclusion is there are many specific methodologies for developing Web and popular Web-based application, in our work is user-based in an approach self-organizing. The use of XML and Java (or C#) languages in connection with RDF formats implements the interoperability among different format of data.

ACKNOWLEDGMENT

The authors wish to thank Ro Yung Jia, and the teacher Dr. Plinio Vilela, of read and comments of this work, Euripides Foundation of Marília, and CAPES-CNPq Brazil to support this project.

REFERENCES

[1] Anderson, J R.. – "Cognitive Psychology and its Implication", W.H. Freeman and Company, New York, 1985.

[2] Braga, A. P.; Ludemir, T. B. and Carvalho, A. C. P. L. F.– "Redes Neurais Artificias – Teoria e Aplicações", LTC Editora, Rio de Janeiro, 2000.

[3] Damásio, A. - O erro de Descartes, Companhia das Letras, São Paulo, Brasil, 1996.

[4] Fausett, 1994 - Fausett, L. – "Fundamentals of Neural Networking – Architecture, Algorithms and Applications", Prentice Hall, Englewood Cliffs, 1994.

[5] Freeman, W.J. – "Three centuries of category errors in studies of the neural basis of consciousness and intentionality", CA, Berkeley, Neural Networks, 10, 1997, pp. 1175-1183

[6] Gibson, .J.J. – "The Ecological Approach to Visual Perception", Boston: Houghton Mifflin, 1979.

[7] Hebb, D. - "A neuropsychological theory. In S. Koch, Ed., *Psychology: A Study of a Science",* 1. New York: McGraw-Hill, 1959.

[8] Heidegger, M. "Being and Time", Trans. MacquarieJ, Robinson E., New York: Harper, 1927;1962.

[9] Honkela, T., Kaski, S. Lagus, K. and Kohonen, T.. "WEBSOM - self-organizing maps of document collections". Proceedings of Workshop on Self-Organizing Maps, Espoo, Finland, June 4-6, Helsinki University of Technology, Neural Networks Research Center, pags. 310-315,1997, url = "citeseer.nj.nec.com/honkela97websom.html".

[10] Jung, Peter, Cornell-Bell, Ann, Madden, Kathleen S. and Moss, Frank "Induced Spiral Waves in Astrocyte Syncytia show self organized Criticality", Journal of Neurophysiology, APStracts, 4:312N, 1997.

[11] Kohonen,. T. Self Organizing Maps. Spring-Verlag, 1995.

[12] Kwok, K. L. - A neural network for probablistic information retrieval. In *Proceedings of the Twelfth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, Cambridge, MA, June 25-28, 1989, pp. 21-30.

[13] Lévy, Pierre – As tecnologias da inteligência, trad. Carlos Irineu da Costa, São Paulo: Ed 34 Ltda., 1999.

[14] Lévy, Pierre – Cibercultura, trad. Carlos Irineu da Costa, Rio de Janeiro: Ed 34 Ltda., 1993.

[15] Lin, X., Soergel, D. and Marchionini., G. – "A self-organizing semantic map for information retrieval"., in: *Proceedings of the Fourteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval*, Chicago, IL, October 13-16 1991, pp. 262-269.

[16] Steel, Chad M. – "Building a Multisite Web Architecture", IEEE Internet Computing, September-November, 2002, pp. 59-66.

[17] Stillings, Neil .et al. - Cognitive Science, an introduction, MIT Press, Cambridge, Massachusetts, 1987.

[18] Tsuda, I. – "Chaotic itinerancy as a dynamical basis of hermeneutics in brain and mind". World Futures, Vol. 32, 1991, pp. 167-184.

[19] Wilkinson, R. and Hingston, P. -"Incorporating the vector space model in a neural network used for document retrieval," Library Hi Tech, vol. 10, no. 1--2, pp. 69--75, 1992.