GCC: A Knowledge Management Environment for Research Centers and Universities

Jonice Oliveira[1], Jano Moreira de Souza[1,2], Rodrigo Miranda[1], Sérgio Rodrigues[1], Viviane Kawamura[1], Rafael Martino[2], Carlos Mello[2], Diogo Krejci[2], Carlos Eduardo Barbosa[2], and Luciano Maia[2]

[1] COPPE/UFRJ - Computer Science Department - Graduate School and Research in Engineering – Federal University of Rio de Janeiro
{jonice, jano, mirand, searo, vka}@cos.ufrj.br
[2] IM/UFRJ - Computer Science Department - Institute of Mathematics – Federal University of Rio de Janeiro (UFRJ)
PO Box: 68.513, ZIP Code: 21.945-970, Rio de Janeiro, Brazil

Abstract. Research centers and universities are knowledge-intensive institutions, where the knowledge creation and distribution are constant – and this knowledge should be managed. In spite of it, scientific work had been known for being solitary work, in which human interaction happened only in small groups within a research domain. Nowadays, due to technology improvements, scientific data from different sources is available, communication between researchers is facilitated and scientific information creation and exchange is faster than in the past. However, the focus on information exchange is too limited to create systems that enable true cooperation and knowledge management in scientific environments. To facilitate a more expressive exchanging, sharing and dissemination of knowledge and its management, we create a scientific knowledge management environment in which researchers may share their data, experiences, ideas, process definition and execution, and obtain all the necessary information to execute their tasks, make decisions, learn and disseminate knowledge.

1 Introduction

People need to interact with other people and to access information to create new knowledge, and this interaction and active seeking, managing, assimilating and exchanging of information has impelled information technology toward new directions. In Science, the need for information increases exponentially, and the knowledge sharing is still not very much impressive. Although we live in the information age, most scientific collaboration relies on face-to-face interactions, paper-centered flow, asynchronous communication, with more expressive interaction made only in small groups.

In business, knowledge has been the force behind millions of strategic and operational decisions, throughout time; however, the recognition of the fact that knowledge is a resource which needs to be managed is relatively recent. Even though universities and research centers are very knowledge-intensive, their decentralized organization, the high complexity of scientific data and information, and peculiar

X. Zhou et al. (Eds.): APWeb 2006, LNCS 3841, pp. 652 – 667, 2006.
© Springer-Verlag Berlin Heidelberg 2006

processes have been obstacles in the move towards more efficient management of scientific knowledge. Other problems occur as a consequence of a low collaboration degree in the scientific environment, as the unknowledgeable about the researchers' competences, and resource waste by the repetition of mistakes and the reinvention of already-known and consolidated solutions.

Some attempts to improve knowledge dissemination have been made in the scientific scenario, and we can mention online scientific journals, the growing use of Internet in universities, e-mail and the adoption of collaboration tools. Nevertheless, information technology must lead to more a fundamental change than automating and accelerating traditional processes [11]. Thus the capabilities of information technology may fundamentally change the way in which scientists work, collaborate, and, consequently, create, organize, and disseminate their knowledge.

Based on these issues, we have developed a web environment the purpose of which is to provide resources to enable the knowledge management in research organizations. Our approach envisions personal knowledge management, process management – allowing the reuse of models and rationale capture, collaboration tools and knowledge visualization and navigation. Special attention was given to identify what knowledge may be present, as the way of when, how, and to whom it could be delivered. Competence management, user profiling and knowledge matching techniques were intensive and they are used to filter the amount of information to be provided. Selective dissemination of information was created to deploy automatically important information to communities. Our approach is based on a national ontology of Science, and provides mechanisms to enrich this one.

The remainder of this paper is organized as follows: Section 2 discusses a number of theoretical Scientific Knowledge aspects and the section 3 explains our scenario, which reflects some common topics with the international scenario. As the objective of this research is proposing an environment to facilitate the collaboration in scientific organizations, knowledge sharing, dissemination and creation, our approach is explained in section 4. The following section (section 5) describes some related work and the difference with our approach. Future works and the conclusion are shown in section 6.

2 Scientific Knowledge

The nature of knowledge has been a matter of intense discussion since the beginning of philosophy [12]. Much research has sought to refine the concept of knowledge and to answer questions about its core characteristics. One of the first to define scientific knowledge was Socrates [6], and, for him, knowing a subject or concept consisted of "gathering the components of a singular thing, or of a real substance, and joining the similar ones, and separating the unsimilar ones, to form the concept or the definition of the singular thing". In this way, in order to "join the similar ones" it is necessary for one to have principles, axioms, definitions and demonstrations, for a concept to be defined as true. In other words, scientific knowledge is the knowledge resulting from scientific activities, and its objective is to demonstrate, by argumentation, a proposed solution to a problem, relative to a certain issue [19]. The most common approach to study knowledge definition, therefore, is to treat the concept as undefined and to approximate its meaning by examining the context of its use [12]. As per [12], an analysis of the use of the term in everyday and scientific language leads to three main interpretations: knowledge-that (objective knowledge), knowledge how (know-how) and knowledge by acquaintance. But, what is the difference between business knowledge and scientific knowledge?

The first difference is related to the analysis of the data used and of the knowledge construction process. Frequently, the activities executed in a business domain are well defined, as well as the knowledge needed for the execution of each of those activities is well known, while the scientific activities comprise sequences of attempts, because the domain is not completely known. In other words, scientific knowledge is built gradually according to the results of a number of activities and it can be subject to constant alterations.

Independent of the complexity of the manipulated data, of the information analyzed and of the way in which is structured, there is another factor in scientific knowledge construction: collaboration. As per [17] collaboration is the essence of science, because there is, on account of people's union, the possibility of knowledge exchange for common activity execution (peer-to-peer collaboration); dissemination of an acquired knowledge (Mentor-Student); researchers with different domains of knowledge who do not share a common background can interact by exchanging results (Interdisciplinary) or just by publishing the research results achieved (Producer-Consumer).

Related to the level of inter-personal collaboration, collaboration and knowledge flow in scientific environments are usually more restricted, and occur among a small number of people working in the same group, dealing with or researching more specific items of the their domain. Many researchers do not know about other researchers who are working with works correlated to theirs, as they are based in a different research center, with the distance hindering regular contact. The use of web-based knowledge management tools is a way to provide better communication and interaction among researchers belonging to a same domain – independent of synchronous communication and physical presence, and then the four types of scientific collaboration proposed by [17] can be applied easily.

3 Our Scenario: Classification, Resources Sharing and Knowledge Loss on Brazilian Scientific Research

Knowledge organization has always been recognized as an area of research and study by professionals of different domains. Computing Science professionals are currently interested in knowledge organization [8]. LANGRIDGE, in [14], emphasizes the fundamental study of classification related with the study of meaning and definition, that is, of semantic.

In knowledge organization, especially in the Science and Technology scenario, we had important contributions such as: classification and indexation in Science [20], in Social Science [10] and Humanities [13]. We have had important progresses in classification theory, as facets Ranganathan [15, 16], concept theory [9, 8] and terminology theory [22;4], and, more recently, principles of ontology construction have been improving the knowledge organization area in the information technology context [5; 3].

Knowledge representation, in some classification structures which permit organizing, systematically, data from published scientific production and other Science activities, is very important for learning, knowledge dissemination, production management and evaluation. In this context, the classification table of "Knowledge Areas", created by CNPq[1], is used of all research centers in Brazil, and appears as support tool. It is a Brazilian attempt for establishing a unique classification, an ontology, at national level, in which academic systems and digital libraries can lean on and use it to classify and organize their information. This classification has some principal areas, as Exact and Earth sciences, Biological Sciences, Engineering, Health Sciences, Agrarian Sciences, Applied Social sciences, Humanities, Linguistics, Languages and Arts. Each area has a sub-tree with its concept categorization.

The main issue is whether a unique classification can be a complex approach of the Science Universe and represents the diversity of involved activities in correlated areas. Satisfying the different interest of institutions on data, information and knowledge aggregation becomes impossible. LANGRIDGE [14] emphasizes that the knowledge unit is a controversial topic, mainly related to knowledge division on disciplines.

The construction of a table of knowledge areas involves basic aspects of organization and classification. The first issue classification pointed by LANGRIDGE [14] mentions that the same objects can be classified in different ways, depending on their purpose. Then, a unique representation can represent wrong, incomplete and inappropriate complex areas of Science. Other critical points of the CNPq classification are i)it is deficient to represent the natural evolution of some areas and how the research grew in the research centers; ii)it does not enable temporal evolution of knowledge areas, taking into account that knowledge areas can be represented in different places in a classification at the time; and, iii)some areas can appear with different names. Then, identifying new areas of Science, capturing temporal changes in knowledge areas and reflecting the production in the research centers is a requisite for any effort to Brazilian scientific production.

Some enterprises and business companies, in partnership with universities and research centers, help with the research development, but most significant and expressive contributions to national scientific research are made by the Government, using national and state agencies. The size and quantity of universities and research centers in Brazil, and the absence of an efficient approach to identify similar interests and projects among these institutions, constitute some problems: a low collaboration degree among scientific organizations, resource waste by the duplication of efforts and the reinvention of already-known and consolidated solutions.

Inside a scientific organization, we have the problem of knowledge loss, which means, institutions lose some experts and specialized professionals, and no attempt of knowledge transmission is made because universities and research centers do not know what they know: their forces (scientific areas with good production) and their weaknesses (scientific areas with insufficient production and few researchers).

---

[1] The National Council for Scientific and Technological Development (CNPq) is a foundation linked to the Ministry of Science and Technology (MCT), to support Brazilian research.

These problems are faced in Brazil, but they are common in several international institutions. Based on this scenario, we developed the GCC. Our approach allows for identifying new knowledge areas, which are not represented in the CNPQ's Knowledge Area classification, from e-meeting logs, projects definition, mental maps, publications, and other ways of user interaction, and having as consequence the improvement of a unique ontology, incorporating new scientific knowledge areas and monitoring the organizational knowledge evolution. This work permits a knowledge management in personal and organization aspects, facilitates the knowledge capture in scientific projects, and helps the intra and inter-institutional collaboration, as will be described in next section.

4 The GCC Architecture

We have envisioned a web-based architecture to enable the knowledge management in scientific environments and increase the collaboration between researchers. This environment is titled GCC, which is the acronym of "Gestão de Conhecimento na COPPE[2]" (Knowledge Management in COPPE). The GCC architecture, as shown in Figure 1, is detailed in the following sections.

The GCC services and main functionalities are:

- **Personal KM Services** – manage users' personal knowledge and data, based on the researchers' "curriculum vitae", weblogs and mental maps.
- **Project Management Services** – manage scientific project execution, enabling the definition of a process, reuse of past processes, and the capture of acquired knowledge in the activities of a process.
- **Community Services** – allow for easy and quick communication, providing tools for synchronous and asynchronous collaboration and dissemination of information and knowledge to communities.
- **Knowledge Visualization and Navigation Services** – displays knowledge and its relationships in a more intuitive and visual way – differently from common reports, allowing the user to interact with the information, navigate and access it.
- **User Profiling and Knowledge Matching Services** – identify researchers' interests, profiles and competence. This service provides information to other modules such as searching for users with similar profiles and whom it might be interesting to establish contact, discovering researchers' competences, suggesting experts to execute a specific activity in a context, and representing their personal interests to a more precise selective-information dissemination.
- **Knowledge Base** – where all kinds of knowledge as processes, past experiences, practices, e-meeting logs, messages exchanged, concept definitions, group and personal characteristics and others are stored.
- **Collaborative Filtering Service** - can streamline research, improve retrieval precision, reduce the amount of time spent looking for significant changes on resources, and even aid in the selection of data, information, people and process definition.

[2] COPPE - Graduate School and Research in Engineering from the Federal University of Rio de Janeiro (UFRJ)-Brazil.

- **Inference Engine** - inference mechanisms that search through the knowledge base and deduce results in an organized manner.
- **Analysis Services** - Our proposal uses tree kinds of analysis services: i) reports of researchers' personal information, process and community status; ii) an OLAP (Online Analytical Processing) structure to provide more solid and evolutionary visions about the researchers' knowledge acquisition process, the knowledge flow in a community and the evolution of concepts (new knowledge created, merger of concepts and knowledge which is not used more); and iii) mechanisms based on Business Intelligence to compare researchers, departments and research centers, arousing the possibility for collaboration.

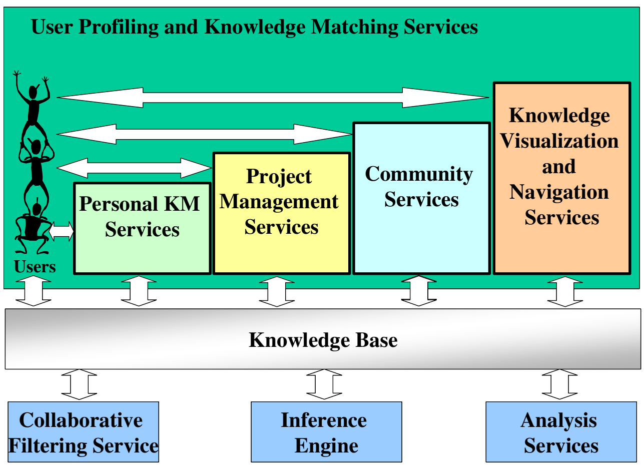

Fig. 1. GCC Architecture

The GCC is a proactive environment, that is, capable of taking initiative according to the researcher's profile and domain, as well as reacting in response to the requests and changes in the environment. In this way, it supplies, at the right time, certain new and relevant knowledge to help researchers in their tasks. The architecture serves to create an effective collaborative and learning environment for all those involved, providing distributed scientific knowledge in a single and accessible system.

In our environment, we use the CNPQ ontology and we enrich this classification with knowledge which flows in the GCC, such as personal knowledge, competences, new knowledge concepts and definition.

The Services are discussed in more detail below.

4.1 Personal KM Services

This module is responsible for providing functions for a researcher to manage his/her personal knowledge, as well as information about him/her. This module provides services such as:

4.1.1 Curriculum Vitae

The curriculum vitae is one way of keeping information about a person. The GCC enables the importation of curricula from CV-Lattes System, a kind of CNPq system. When the user does not have, or does not want use, the curriculum in CV-Lattes System, he/she can fill out his/her information in the GCC. This information comprises the name, personal information such as address, e-mail, home page and phone, academic background, professional activities, language skills, scientific production, advisory and prices. In addition to this information, the user should say in which Knowledge Areas he/she acts and what his/her competences and his/her degree of expertise are. Competences are abilities and knowledge areas, which are not represented in the CNPq classification, in which the user works and has some fluency. The user can also show some areas in which he/she has some interest, but in which he/she is not an expert yet.

4.1.2 Personal Blog

Weblogs may be viewed as personal Web pages or "home pages". The term refers to a web site that is a "log of the Web", indicating a record that points to material available on the World Wide Web. Many say that weblogs comprise an electronic-diary, but in the GCC this acts as a tool to provide personal knowledge management. Weblogs, in general, have a number of features which are as follows:

– **Personal editorship** - The content of the site is under the responsibility of a single person. In our case, the user is a researcher, a GCC user, and the weblog reflects some topics about his/her individual's profile.
– **Hyperlinked post structure** - The weblog's contents consists of typically short posts that feature hypertext links referencing material outside the site. The selection of links is entirely up to the editor, who may link anywhere on the web. There is also no prescribed length for a post - some posts simply consist of a single link to content elsewhere, but most often they also include additional information and/or personal commentary on the issue under discussion. In the GCC this may be information about successful or unsuccessful experiments, lesson notes, and other kinds of scientific information.
– **A first pass before a community creation** - An enormous amount of content is published daily in the Web. As it is impossible to read it, people need the means of filtering this output to find the material that will be most relevant to them. A weblog operates in much the same manner. By reading a weblog that is edited by someone with interests similar to yours, you obtain a view of possibly relevant material. In the GCC, by combining the output of several chosen weblogs, you obtain a tailor-made publication and contact with researchers with the same interest as yours.
– **Frequent updates** – which are displayed in reverse chronological order can show in GCC the researcher-interest evolution and involvement about a topic, as a chronological record of your thoughts, knowledge, references and other notes that could otherwise be lost or disorganized.

More than the issues cited above, a weblog is an important tool in personal knowledge management because it embodies several important functions for the researcher who uses them, such as:

In the GCC each researcher can have a weblog, and pieces of text can be turned private (only the author sees the description). By means of “User Profiling and Knowledge Matching Services” tools, we can use the weblogs to automatically identify: a researcher’s personal interests, his/her competences, and knowledge areas cited, and infer which areas are connected and how they are connected.

4.1.3 Mental Maps

Knowledge representation as mental maps, in which the concepts are organized in classes and sub-classes, is a way to structure the information. Since every one of us creatively constructs our own maps, then these will be different from everyone else's maps. Each of us will have different perceptions of our needs, different learning styles, and even perceived shared experiences differently.

In the GCC, the researcher can construct mental maps to define concepts, help in brainstorm sessions and simplify the job of discussion in a graphic way (Figure 2). Users can display some concepts as public, so that anyone can see the concept, its relationship with other public concepts, and its definition. In addition to the information organizations, mental maps have important function in learning in the GCC.

Our own maps of reality (and not reality itself) determine how we interpret and react to the world around us and give meaning to our experiences and behaviors. No individual map is any more 'true' or 'real' than any other. The wisest and most compassionate maps are those which make available the wisest and richest number of choices, as opposed to being the most 'real' or 'accurate'. This has implications when we are identifying outcomes, planning learning activities, assessing learning: gaining the most effective result is a consequence of viewing these issues from multiple perspectives. Exchanging maps, analyzing the different representations of reality, collaborating, discussing and trying an agreement about a definition allow people to discover new definitions, learn and enrich the domain.

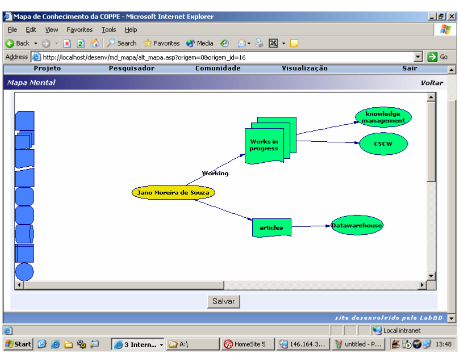

Fig. 2. Mental Map Editor

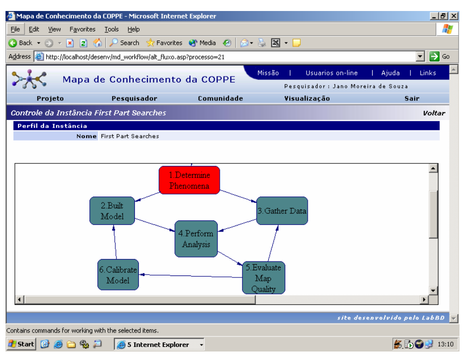

Fig. 3. Workflow Editor

4.2 Project Management Services

This module is responsible for providing services to: i) define and execute a project by a workflow tool, ii) store the knowledge created during project execution, iii) permit knowledge reuse.

4.2.1 Define and Execute a Project

The project manager or responsible researcher creates the project model, with the activity sequence and the needs for the execution of each activity: competences necessary for execution, inputs, outputs and tools. The project-model creation is defined in a workflow graphic tool (Figure 3), and process execution control is made by a workflow engine. Finding a person with a specific competence to execute an activity needs more semantic attention because if nobody with the desired competence is found, we can recommend people with competences similar to the competence needed. This search is a functionality provided by the "User Profiling and Knowledge Matching Services". This service recommends people who better match the pre-requisite competence in an activity and, consequently, can perform this activity better.

4.2.2 Knowledge Storage During the Project

During activity execution, the involved parties can add all kinds of explicit knowledge to this activity, such as reports, documents, comments, suggestions, better practices, mistakes and ideas. Then all steps in the activity execution can be documented, and in the future we can track all kinds of created knowledge and the context it happens, important for an inexperienced researcher to learn about the process in a domain. The process model itself is a kind of knowledge which should be stored for further access.

4.2.3 Knowledge Reuse

Previous process-models and the knowledge described in their activity can be reuses totally or partially by other users.

4.3 Community Services

One of the main focus in the GCC is virtual-community creation, created by groups of researchers with a common interest, who could exchange information and work together. This module displays some tools to improve the interaction between people in a community, such as:

- **Survey** - A question is created by the community supervisor and answered by all members. The presentation of the survey is random and the questioner can change the priority of the queries. The application still provides a result with a graph of inquires and their response percentage.
- **Forum** - To support posted messages, a forum service for communities was built for asynchronous communication.
- **News** – Any member of a community can send news about related topics, such as links and external materials which underlie it.
- **Scheduled E-meetings** – A member can ask for a private and synchronous e-meeting with another member, as in the case of an inexperienced researcher asking for an explanation of a doubt from an expert researcher. Then, the member invites

another member for an e-meeting. If the invited member agrees the meeting is confirmed and appointed in the community schedule.

- **Public library** – A space to display electronic material and links.
- **Log Categorization and Storage** – All conversation of an e-meeting is stored and categorized in one or more Knowledge Areas. This is important because a conversation or interview is a kind of explicit knowledge which could be consulted in the future, avoiding future interviews and accelerating the knowledge capture.
- **Community Evaluation** – Several metrics are used to control the community evaluation and the member interaction. All evaluations are public.
- **Events** – Several events such as conferences, workshops, lectures, etc, can be posted here.

4.4 Knowledge Visualization and Navigation Services

The typical visualization in the majority systems purport a litany of complaints including unfriendly interfaces, the absence of intuitive search structures, and the requirement that users learn special languages or conventions in order to interact effectively with the online systems. Based on these complains, more than the reports, lists and graphics visualization provided by the GCC, the environment has two more kinds of knowledge visualization and navigation services: the Hyperbolic Tree and the Conceptual Project Map.

4.4.1 Hyperbolic Tree

This is a hierarchical tree structure to a hyperbolic display citing two significant qualities of the structure: i)the nodes or components of the tree diminish in size the farther away they are from the center of the display and ii)the number of nodes or components grows exponentially from parent to child. In the GCC, as shown in Figure 4, the hyperbolic tree is used to visualize the CNPq's Knowledge Area Classification. The user can navigate by the CNPQ classification, as shown in Figure 5, and by all information related to a knowledge area, as projects (in green), competences, people (in blue), communities, and etc.

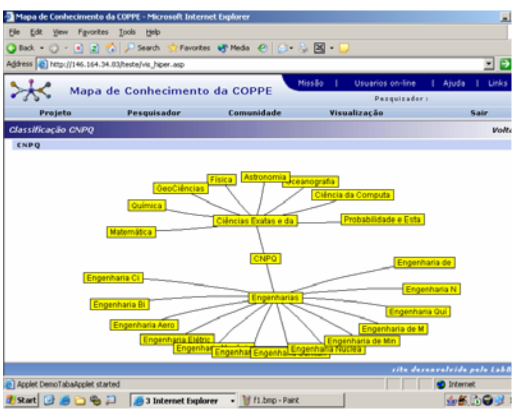

Fig. 4. Hyperbolic Tree of CNPq's Knowledge Areas

Fig. 5. Navigation in Hyperbolic Tree

4.4.2 Conceptual Project Map

Permits a tree-like visualization of a specific project, as Figure 6. Then the user can see all related information of a project, such as materials, contributions, participants, collaborators, managers, knowledge areas and competences which are pre-requisites and the workflow model. This map permits searching over it, and opening a node which matches.

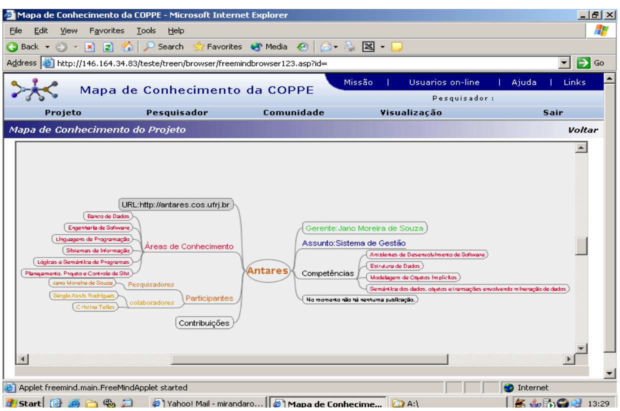

Fig. 6. Project Conceptual Map

4.5 User Profiling and Knowledge Matching Services

Nowadays, the main problem in scientific organization is the inability to discover what it knows, what is its abilities are, what researchers and experts know and in which knowledge areas they work. Identifying the researchers' profiles, their interests and expertise enhances the chances of collaboration. The GCC provides three kinds of services: i)the S-Miner, to identify automatically the competence of a researcher, using his/her publications; ii) the Competence Search, a semantic search for competences and iii)Web Miner, a way of tracking pages the user accesses when navigating the Internet and identifying topics of his/her interest.

4.5.1 S-Miner

Fundamentally, the S-Miner function is to mine competencies based on researchers' publications'. It is used to extract key words derived from these publications, contrast with the other publications histories and suggest, to the text owner, the feasible communities rooted in the mapped competencies. Initially, the text is submitted to the tokenization algorithm. Tokenization purports in the word (tokens) identification. After breaking of the text in tokens, the process continues with the elimination of insignificant words – named Stop Words. The collection of Stop Words which will be cleared away from the text is called Stop List. This catalog of irrelevant words is strongly dependent on the language and the applied circumstance – the S-Miner can handle the English and Portuguese (Brazilian) languages. When Stop Words are removed, the remaining words are considered filtered and should then enter a new selection process. In this phase, the next procedure comprises the creation of weights for each word type. We use the Stemming technique to measure the relevance of a term by removing suffixes in an automatic operation. Ignoring the issue where the words are precisely originated, we can say that a document is represented by a vector of words, or terms. Terms with a common stem will usually have similar meanings, for example: CONNECT, CONNECTED, CONNECTING, CONNECTION, CONNECTIONS. Then, after having the words, stemming to count the relevance of a term is applied. Terms are related to a person's competence, and these competences and the degree of expertise (using the measure of relevance of a term) are stored in database.

Besides, these filtered words, named Relevant Words, are submitted to the association between competences and words. This association contemplates the competences established in CNPq Knowledge Tree and the Relevant Words. This connection suggests that each competence of the CNPq Knowledge Tree can be derived from a set of key words. This relationship is restored from a dictionary stored in the GCC. Finally, after the mining, it is possible to verify the mapped abilities in a report provided by the application.

4.5.2 Competence Search

This is a service which provides a list of people with specific competence to perform an activity of a process or to lead a community. The searcher architecture is based on the GCC database and the SMiner provided mining. Thus, the competences are searched following this order of priority:

- **Declared competences** – the competences which the researcher declares to have in GCC or Curriculum Lattes.
- **Project competences** – correspond the competences which were the pre-requisite of a project. We imagine researchers who worked in a project, known about the competence needed.
- **Extracted competences** – recovered from the researchers' published text mining by S-Miner.
- **Community competences** – collected from the communities in which the researcher participates or contributes.

In addition, the Competence Searcher includes distinctive weights for each type of competence found: Declared competences- weight 3, Project competences- weight 2, Community competences- weight 1. Consequently, the minimal weight was chosen not to misrepresent the analysis.

While the competences above are gathered, the Competence Searcher identifies the levels of each kind of competence in the CNPq Knowledge Tree. After discovering it, it is viable to achieve the researcher relevance rooted on the searched competences and, after this process, a list of researchers who bear the searched abilities. On the other hand, if the Searcher does not find the exact level, that is, if some competence is not correlated directly in the CNPq Knowledge Tree, the Competence Searcher will go up the tree, searching the levels up to the root. In this case, there is no accumulation of weights, since this action is only used in the absence of competence linked straight.

4.5.3 Web Miner

A way of tracking the pages the user accesses when navigating the Internet and identifying topics of his/her interest by analyzing the log web. The user can rate sites according to their relevance to his/her interest. This information is inferred from time spent on sites, number of accesses, re-visiting. Sites can be recommended to other users with similar interests.

4.6 Collaborative Filter

In this service, documents, people and process' models are recommended to a new user based on the stated preferences of other, similar users and communities. We use a Collaborative Filtering Spider which collects, from the Web, lists of semantically related documents and disposed to communities.

4.7 Inference Engine

Reasoning about profiling is made in the "User Profiling and Knowledge Matching Services". This module is responsible for extending these functionalities to infer process, cases or solutions, and reusable content.

4.8 Analyses Services

We propose some tools for observing the knowledge evolution, the evolution of communities and comparative analysis by researchers, departments and organizations. These kinds of analysis are especially useful to:

- The intellectual capital of the institution not be associated exclusively to people who own critical knowledge, but can be distributed among the members of a research team;
- Make the identification of knowledge areas with a shortage of professional possible and then plan a way to acquire this knowledge, by training or external researcher recruiting;
- Make the regular appraisal of each researcher's knowledge level possible;
- Analyze chances of external collaboration;
- Analyze the quality of an institution; and
- Identify the appearance, death and merging of knowledge areas

Our proposal uses an OLAP (Online Analytical Processing) structure to provide more solid and evolutionary visions about the researchers' knowledge acquisition process and knowledge flow in a community. Moreover, GCC uses techniques of Business Intelligence to the other issues.

5 Related Work

Some work the purpose of which is to improve knowledge sharing and collaboration among researchers has been created, although focus was different as in:

- The Pacific Northwest National Laboratory (PNNL) [7], is developing and integrating an environment of collaborative tools, the CORE (Collaborative Research Environment) [17].
- The CLARE [21] is a CSCL environment (Computer-Supported Cooperative Learning) which aims at scientific knowledge creation by cooperative learning. For this, CLARE uses a semiformal representation language called RESRA and a process model called SECAI.
- The Science Desk Project[18] is developed by NASA and its purpose is the development of a computational infrastructure and a suite of collaborative tools to support the day-to-day activities of science teams.

Moreover, projects and some methodologies were created to aim the construction of Knowledge Management Systems, the ISKM ("Integrated Systems for Knowledge Management") [1;2], an approach to aiming at community creation, managing the necessary knowledge in a learning environment and decision-making applied to the natural resource management domain.

Our approach differs significantly from the above mentioned because major attention is given to competence management, collaborative filtering, knowledge navigation and user profiling and knowledge-matching strategies, and personal knowledge management, as well. Special attention is given to competence management, because experts and their knowledge are the most important assets in a scientific institution. Moreover, the GCC makes it possible for researchers to collaborate and interact among themselves, facilitating the communication between people within the same research area, gathering, in one environment, different perspectives and expertise present in the organization, enabling the formation and recognition of groups with common interests, diminishing the amount of time spent in the coordination of team work and expediting the project problem solution processes. As explained before, the GCC also proposes an analysis structure providing evolutionary visualization of the knowledge buildup and the comparison to common interest matching among departments and organizations to allow for possible collaboration. None of these related works attempts to automatically identify new knowledge areas and correlate them.

6 Conclusion and Future Work

Our work is a scientific knowledge management environment, which aims to encourage knowledge dissemination, expert localization and collaboration. We also envision knowledge reuse and the detection of new areas of knowledge in Science, thereby enriching Science Ontology used in Brazil.

One way of identifying new knowledge areas in Science is monitoring current scientific activities, as publications, projects, groups and the individual researcher. Based on this, in addition functionalities aiming at scientific work, we automatically identify knowledge areas in order to: i)improve communication among people, ii)identify people to execute scientific activities, iii)know what the institution knows, iv) reuse knowledge and v)update the national Science ontology, and detect new emerging knowledge areas.

Currently, the GCC is a centralized-base environment. As future work, we envision extending it to a distributed scenario. Moreover, we will measure how much the GCC is aiming at those researching in order to manage their personal knowledge, disseminate it, their learning and collaborating.

This approach is in an evaluation process and it is been used by the Database Group of COPPE[2]. Our future work is extend and makes available to all pubic universities and research centers in Brazil.

References

1. ALLEN, W., 2003, "ISKM (Integrated Systems for Knowledge Management)". In: http://nrm.massey.ac.nz/changelinks/iskm.html, Accessed in 21/05/2002.
2. ALLEN, W.,KILVINGTON, M., 2003, "ISKM (Integrated Systems for Knowledge Management): An outline of a participatory approach to environmental research and developement initiatives". In: http://www.landcareresearch.co.nz/sal/iskm.html, Accessed in 21/05/2002.
3. BIOLCHINI, J. C. de A.. Da Organização do Conhecimento à Inteligência: o desenvolv imento de ontologias como suporte à decisão médica. Ph.D. Thesis. Rio de Janeiro, 2003.
4. BRANSFORD, J. D., BROWN, A.L. and COCKING, R.. How people learn: Brain, mind, experience and school. Washington, D.C.: National Academy Press, p. 10, 1999
5. CAMPOS, M.L. de A. A organização de unidades de conhecimento em hiperdocumentos: o modelo conceitual como um espaço comunicacional para a realização da autoria. Ph.D. Thesis. Rio de Janeiro, 2001.187p.
6. CHAUI, M.. Introdução à história da filosofia: dos pré-socráticos a Aristóteles. 2a.edition, São Paulo-Brazil, 2002 – in Portuguese.
7. CHIN, G., LEUNG, L. R., SCHUCHARDT, K., et al, 2002, "New Paradigms in Problem Solving Environments for Scientific Computing",Anais do IUI'02 - Association for Computing Machinery (ACM) San Francisco, California, USA
8. DAHLBERG, I. Faceted classification and terminology. In: TKE'93, TERMINOLOGY AND KNOWLEDGE ENGENEERING, Cologne, Aug.25-27, 1993. Proceedings. Frankfurt, Indeks Verlag, 1993. P. 225-234
9. DAHLBERG, I. Ontical structures and universal classification. Bangalore, Sarada Ranganathan Endowment, 1978 a . 64 p.
10. FOSKETT, D.J. Classification and indexing in the social sciences. London, Butterworths, 1963.190p.
11. HAMMER, M. and CHAMPY, J.. Reengineering the Corporation. Harper Business, 1994.
12. HARS, A. From publishing to knowledge networks : reinventing online knowledge infrast ructures. Berlin ; New York : Springer, 2003.
13. LANGRIDGE, D. Classification and indexing in the humanities. London, Butterworths, 1976.143p.

14. LANGRIDGE, D. W. Classificação: abordagem para estudantes de biblioteconomia. Tradução de Rosali P. Fernandez. Rio de Janeiro, Interciência, 1977.
15. 15. RANGANATHAN, S.R. Prolegomena to library classification . Bombay, Asia Publishing House, 1967. 640 p.
16. RANGANATHAN, S.R. Philosophy of library classification. New Delhi, Ejnar Munks gaard, 1951.
17. SCHUR, A., KEATING, K.A., et al . Experiment-Oriented Scientific Collaborative Suites for Experiment-Oriented Scientific Research, ACM Interactions, Vol 5, Issue 3, pp. 40-47, May/June 1998.
18. SCIENCEDESK, 2002, "ScienceDesk Project Overview". In: http://ic.arc.nasa.gov/publications/pdf/2000-0199.pdf, Accessed in 05/2002.
19. SEVERINO, A. J..Metodologia do Trabalho Científico. 22a.edition. Rio de Janeiro – Brazil, 2002– in Portuguese.
20. VICKERY, B. C. Classification and indexing in science. London, Butterworths, 1975.
21. Wan, Dadong and Johnson, Philip M., 1994, "Computer Supported Collaborative Learning Using CLARE: The Approach and Experimental Findings", ACM Computing Surveys, p. 187-198
22. WÜESTER, E. L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. In: RONDEAU, G., FELDER, F. (org.). Textes choisis de terminologie I. Fondéments théoriques de la terminologie. Quebéc, GIRSTERM,1981. P. 57-114.