A multiagent knowledge and information network approach for managing research assets

Francisco J. Cantú, Héctor G. Ceballos
fcantu@itesm.mx, ceballos@itesm.mx

Tecnológico de Monterrey
Avenida Eugenio Garza Sada No. 2501, Monterrey N.L., México, 64849

Abstract

We describe a knowledge and information network approach for managing research assets in a knowledge oriented organization using a multiagent system. The purpose of the approach is to provide decision makers a knowledge management framework to assist them in generating benefits from the knowledge assets generated by the research groups in a knowledge institution. Research assets under consideration are of three types: research products, intellectual capital and research programs. Research products include publications of various kinds such as journal articles, research books, patents, technology licensing, trademarks, incubation of technology-based start-up companies and other publications. Intellectual capital consists of the talent and expertise of research staff like professors, students and researchers. Research programs consist of academic curricula, research units, research infrastructure, and business incubators. The approach is supported by an intelligent platform that contains an information system, a multi-agent-based system, a knowledge management system and a knowledge-information interpreter that coordinate repositories, domain ontologies and databases for handling the various types of research assets. The system provides a means of distribution of existing research assets both within the organization and abroad, a variety of research reports, on-line consultations, a search engine, web services and data mining facilities for knowledge extraction. The knowledge and information generated by the system guides managers in defining strategies on competitiveness such as rankings, benchmarking, intellectual property and incubation of technology-based spin-offs. The system has been operational at Tecnológico de Monterrey since August 2004, and has proved useful for both, acquiring consciousness of the research capabilities and for stimulating entrepreneurial science initiatives.

Keywords

Knowledge-based Systems, Multiagent Systems, Knowledge Management, Research Assets, Ontology, Entrepreneurial Science

1. INTRODUCTION

Knowledge has become a strategic weapon in modern society. The level of development and productivity of countries, the income per capita of their citizens and the population welfare are measured in terms of the level of investment in science and technology committed by their government. Generating knowledge is a necessity in modern economies whereby the most important aspect is the economic value that can be attained from the knowledge assets created by key payers of knowledge economies. The economic value is often measured by the number of technology-based companies, the number of jobs that require personnel with high-level training and skills, and the total revenue of those companies as a percentage of the Gross National Product. The universities of the XXI century play an important role as they become engines of economic development for the society (Etzkowitz and Leydesdorff, 2001).

Managing the knowledge generated by an organization and obtaining benefits from that knowledge poses important challenges for any institution. This is true of knowledge-based organizations such as research universities and technology-based corporations. Scientific knowledge is generated mainly in research centers, institutes and academic departments of universities, companies and government-supported laboratories. This knowledge is socially validated by the scientific community after a peer-review process and made public by its publication in journal articles, conference papers, technical reports, magazine articles and other means. On the other hand, the inventions and innovations are kept secret until an intellectual protection is obtained by means of a patent, a trade-mark, a distinctive sign, industrial secrets or other intellectual property mechanisms. The scientific productivity of an organization as well as of its researchers is commonly measured in terms of the quality of the journals in which they publish. Quality is determined by the indexes and impact factor of the journal that depends on the number of cites made by articles in other journals. For inventions and innovations the measure is rather its economic impact that depends on licensing, spin-offs and other financial measures. Combining both types of knowledge seems to be a challenge for knowledge-based organizations. In this research, we focus on the management of scientific and innovation knowledge in organization such as universities and technology-based companies (Alvarado et al, 2007). The main trust is generating benefits from the knowledge assets generated by scientific research and development.

The integration of existing knowledge and information technologies to manage in novel ways knowledge in modern organizations is a key aspect of this article which is organized as follows: Section 2 presents background information; section 3 describes the knowledge and information network approach. Section 4 describes the computer and multiagent-based system design and implementation; section 5 presents a case study at Tecnológico de Monterrey which has designed and deployed the knowledge and information network approach for managing its research assets and obtaining both economic and intangible benefits from them. Finally, section 6 presents the conclusions and future work.

2. BACKGROUND AND RELATED WORK

Knowledge-based organizations build their competitiveness from the market share of products and services they offer and for the value perceived in its products by its potential customers. Often, this value comes from technology research and innovation areas of the organization which are shaped by design and engineering teams for its implementation followed by a marketing plan for launching new or improved products or services into the market. This is also true for organizations such as colleges and universities which emphasize research and development as part of its mission statement. For instance, universities and graduate schools worldwide are annually ranked by various agencies using various indicators. These rakings often have a media impact and are used for student enrollment as well as for obtaining research grants. Research evaluation is an important component of those rankings that are calculated from weighted indicators that yield a general score. Typical indicators are the quality of journal articles and number of citations, the faculty awards, the education of graduate students and researchers, research expenditure, and nowadays, the number of start-ups companies generated by the research and development activities of the professors and the students. Research evaluation has created the necessity for information systems that assist officers in integrating data, information and knowledge for the calculation of these indicators for strategic planning, investment analysis and competitiveness. For instance, America's Best Graduate Schools (Zuckerman, 2004) is based on expert opinion about program quality and statistical indicators that measure the quality of school's faculty, research and students. Academic Ranking of World Universities[1] is based on several indicators of academic or research performance, including alumni and staff winning Nobel Prizes and

[1] http://ed.sjtu.edu.cn/ranking.htm

Fields Medals, highly cited researchers, articles published in Nature and Science, articles in Science Citation Index-expanded and Social Science Citation Index, and academic performance with respect to the size of an institution. The TIMES Higher Education's World University Ranking ranks universities according to peer review, employer review, teaching quality, research quality and international outlook (O´Leary et al, 2009). Indexing services of various companies provide valuable statistics about article and patent citations, abstracts and impact factors. Thomson's ISI Web of Science[2], for example, keeps proprietary databases of papers and patents published on a selected set of journals and conferences; articles and journals are ranked through an impact factor based on the number of citations. Google Scholar[3] provide this service for free obtaining their information through automatic web data extraction. Patents are available through public databases such as the US Trade Patent Office[4]. This creates the need for the novel integration of automated aids in managing the vast amount of data, information and knowledge generated by the network of research activities, which is the main issue addressed in this article.

Knowledge Management studies the processes around knowledge such as its creation, storage, distribution and use (Liebowitz, 2004). In our approach, all this information is managed by a knowledge-based organization using KM methodologies like the one proposed by Liebowitz (Liebowitz and Beckman, 1998). The KM methodology comprises the following steps: Identify, Collect, Select, Store, Share, Apply, Create and Sell. Each of the processes summarized by these verbs is related to the creation of a corporate memory which stores and distributes information and knowledge that is relevant for business operation. They classify corporate memories according to the *capture* and *distribution* mechanisms that are used to build the memory; these mechanisms can be either *passive* or *active* procedures. When capture and distribution are both active the corporate memory is called a knowledge pump. The knowledge and information

[2] http://www.thomson.com/

[3] http://scholar.google.com/

[4] http://www.uspto.gov/

knowledge for managing research assets described in this article corresponds to a knowledge pump.

All these operations are usually performed by staff in the organization but as the volume of information increases, the need for automating these processes becomes important. In this sense, the notion of intelligent agent is a useful tool for the automation of specialized knowledge intensive tasks. Agents are regarded intelligent if they pursue their goals and optimize its performance according some metrics. An agent should be capable of interacting with other agents (human or software) and are designed in such a way that the system of which they are part, achieve a set of global objectives through the interaction of the various types of agents. An arrangement of agents is called a Multiagent Systems (MAS) (Weiss, 1999). There are several methodologies for developing MAS, one of them GAIA (Wooldridge et al, 2000), based on KM methodologies specialized on representing the knowledge of a human expert, and Prometheus (Padgham and Winikoff, 2004), which proposes a practical approach for passing from system specification to agent detailed design. The Electronic Institutions formalism has been proposed as a framework for regulating interactions among heterogeneous agents in open systems. Participant agents modeled through roles must perform according to a specification given in terms of a dialogical framework. Agent roles can be internal, representing autonomous agents, or external, representing human users. An illocution mechanism is used for information exchange and interaction for both kind of agents, software and human (Sierra and Noriega, 1997).

Description Logics (DL) are a family of knowledge representation languages for defining classes of entities and relationships among them. DL derives from semantic networks and frameworks used in the 80s for representing knowledge. Ontology is a kind of dictionary that defines the semantics of terms of a problem domain and is used by agents to reason about domain situations (Uschold and Gruninger, 1996). Ontology Web Language (OWL) is a standard recommended by the World Wide Web Consortium (W3C) for defining ontologies built by using DLs.

Mapping data and information into knowledge and vice versa is an important component of knowledge-based systems. Information obtained from data by data processing mechanisms is well understood in business information systems whereas converting data and information into knowledge by reasoning, pattern recognition and data mining mechanisms has become frequent in knowledge-based systems. Object-relational mapping (ORM) is a programming technique for converting data from relational databases into classes of objects in object-oriented programming[5]. Similarly, as semantic web technologies are maturing there is a growing need for semantic enabled applications based on standards such as the resource description frameworks (RDF) for accessing vast amounts of data in web sites and databases[6].

The integration of existing knowledge and information technologies to manage in novel ways knowledge in modern organizations, such as scientific and technical knowledge, is then the challenge addressed in this article.

3. KNOWLEDGE AND INFORMATION NETWORK

We describe a knowledge and information network (KIN) approach for managing assets generated by scientific research in a knowledge organization. KIN is a kind of social network of humans and intelligent agents. The KIN approach consists both of a web of entities and relationships, and of a multiagent system that coordinates the tasks performed by these entities. The entities are knowledge objects of various types depending on the problem domain that reside in knowledge repositories, databases, web sites and data warehouses. The relationships are access paths among repositories to answer queries or to build reports requested by either human or software agents. The multiagent system sustains an infrastructure in which governance and flow control knowledge is maintained for handling the complexity of the interactions among entities at various levels of design and implementation. Repositories are found at both extra-net and

[5] http://en.wikipedia.org/wiki/Object-relational_mapping

[6] http://www4.wiwiss.fu-berlin.de/bizer/D2RQ/spec/

intra-net levels. The extranet level repositories comprise public databases, web sites of documents and scientific publications available on internet. Proprietary databases such as the Web of Knowledge and SCOPUS are also at this level. The intranet level repositories consist of a corporate memory with various types of databases, portals and web sites. Figure 1 depicts the KIN and its context.

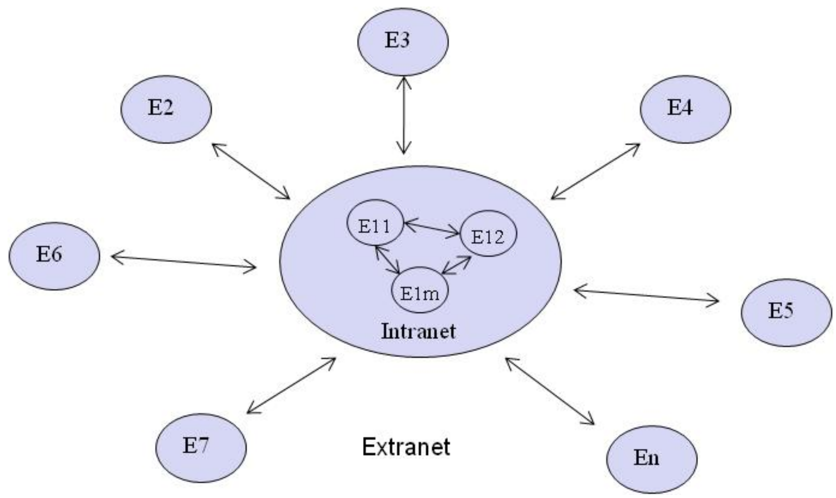


Figure 1. Knowledge and information network with internal and external elements

3.1 Network Components

The knowledge and information network contains three types of components: the research products, the human resources or intellectual capital and the research programs. Research assets include the following: publications of various types such as journal articles, conference papers, books, book chapters, theses, technical reports, white papers, etc. and inventions, including patents, licensing, trademarks, utility models, and other intellectual property elements. Human resources are a key element of any knowledge organization, and these are mainly researchers -faculty members, research assistants, postdoctoral positions and graduate students as well as members of staff, both technical and administrative. The research programs are of different sorts; for a university it contemplates graduate curricula and company incubators, and for other knowledge institutions, it comprises research units, laboratories and infrastructure. All of these elements interconnect with each other establishing a web of relationships which constitute a network through which data, information and knowledge flow to assist decision makers with statistics, trends and indicators of the organization performance. The network components are displayed in figure 2.

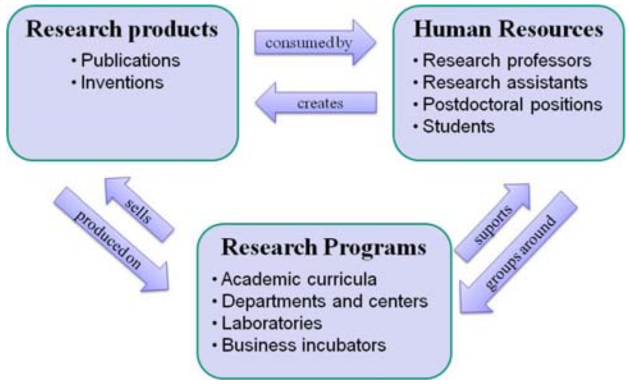


Figure 2. Conceptual model of the knowledge and information network

Each network component is on its own a system comprising a set of knowledge repositories and intelligent software agents. For instance, the research products consist of a repository of publications which follows the taxonomy shown in figure 3. Publications may be refereed or not, if refereed, they may be specialized or not, and if they are specialized publications they may be articles appearing in indexed journals or in conference papers. Some publications are inventions like patents, utility models, industrial designs, industrial secrets, trademarks, or technical reports. Spread out papers like white papers may have less strict review process as well as the various types of books shown in the figure 3.

The value associated to a publication depends on the type of publication. We identify scientific value as well as economic value. Journal articles are assigned the highest score. This depends on the prestige of the journal that is determined by its *impact factor*. The impact factor depends on the number of cites the articles receive from articles of other journals (Garfield, 1999). The ranking of universities as well as faculty awards strongly depend on this measure. On the other hand, the economic value of a publication such as a patent is dependent upon the royalties and revenues obtained as a result of the licensing of technologies and products developed around those inventions for commercial purposes. Universities in the XXI century face the challenge of making these two views of publications coexist in a sustainable way (Etzkowitz, 2001).

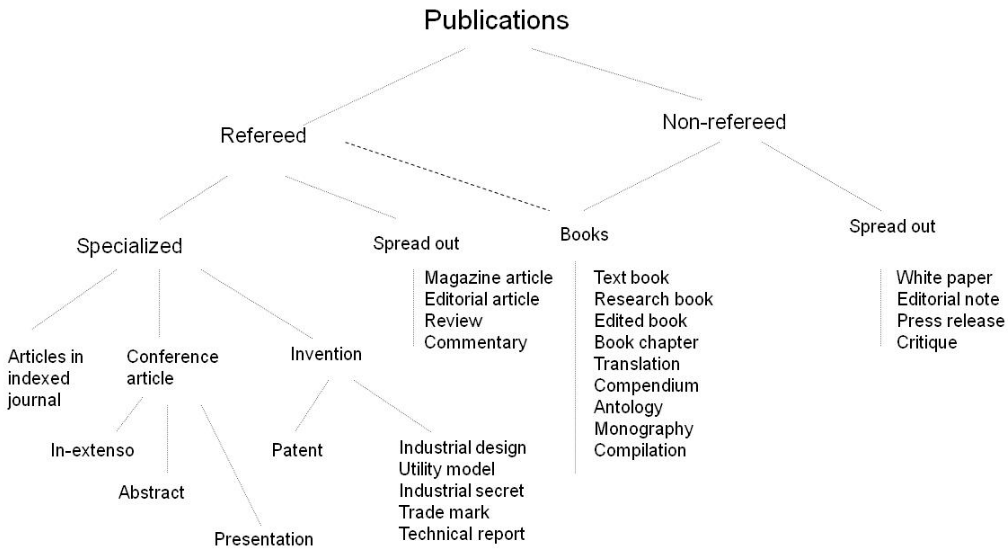


Figure 3. Research products: publications taxonomy

Human resources constitute the intellectual capital of any knowledge organization. In our approach, the key elements are the persons that generate knowledge, namely researchers. Researchers may include research professors, full time researchers such as postdoctoral researchers or research assistants holding PhD degrees, as well as research students of various levels. Research staff consists of officers holding managerial positions including provost, deans or chairmen, technicians at laboratories or experimental sites, and administrative staff. Figure 4 shows the taxonomy.

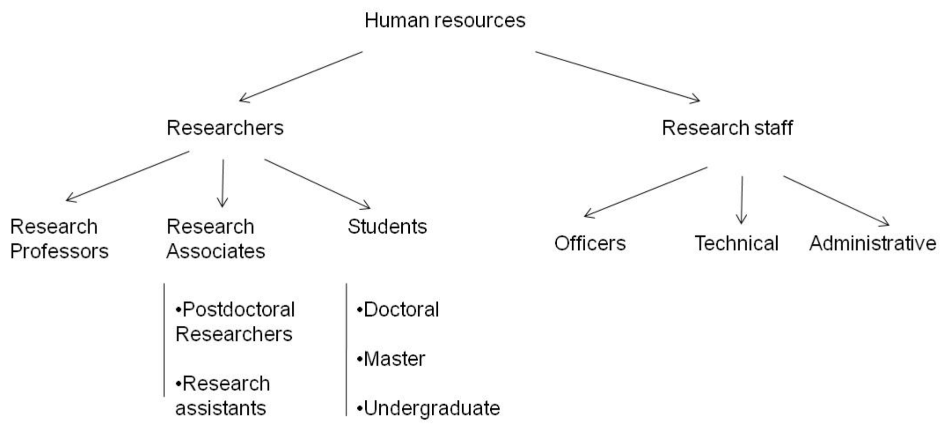


Figure 4. Intellectual capital: human resources taxonomy

The research program components are the organizational units that facilitate the research processes. Among these are academic curricula at the graduate and undergraduate levels; research units such as departments, centers, and infrastructure such as laboratories, experimental units and technology-based business incubators; and research and development projects sponsored by internal funding or external agencies, with a set of deliverables and performed by research and innovation personnel. Figure 5 displays the taxonomy.

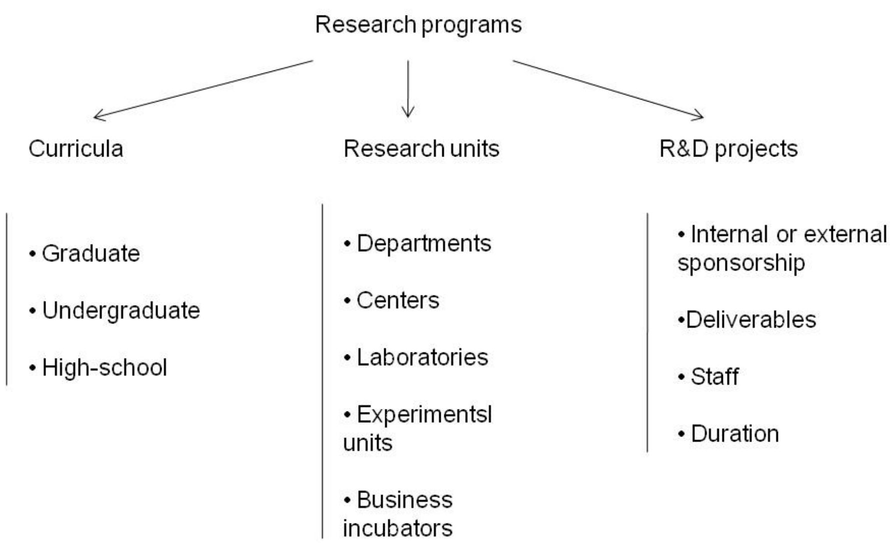


Figure 5. Research programs taxonomy

3.2 Knowledge representation

Knowledge and information in the network is stored in a corporate memory that uses various representation formalisms. KIN entities are defined by ontology formalism and its content is supported by knowledge repositories whose access is regulated by a set of software agents. For instance, the publications repository stores their attributes and the document files. A journal article would have the title, authors, journal, dates, citations, impact factors and other features of the article. For a patent, there is title, inventors, status, countries, licensing contracts, clients, technology claims and other attributes. The same is true for theses and for other types of publications. It is possible to classify the documents in the repository, obtain statistics and do text mining on them. Regarding human resources repositories, research professors are described by a set of attributes with all the information necessary to generate the curriculum vitae of the professor. The same is true of researchers, and research staff. For research students, there is a track of their progress and performance and the student repository stores their transcripts, theses, projects and publications for both current students and those that are now alumni. This is a kind of registrar at research level. With respect to research programs, the taxonomy is also supported by ontology and a knowledge repository in which the attributes of each type of program is formally defined and stored. For an academic curriculum there is a name, director, course outline, past and current students, faculty that teach it, and statistics of the program performance in terms of graduation rates, accreditations, rankings and other features. The same is true of research units like a department, a center or a business incubator. For laboratories and experimental units, the equipment available, its capacity, maintenance and policies for its use are part of the repository.

3.3 Knowledge and information network users

Knowledge and information users are of various kinds. Among the internal users are professors, students and research staff as well as the officers that hold executive or managerial roles within the various hierarchical levels of the organization. The KIN system learns those roles and defines views of the corporate memory according to hierarchical level in the organization. For instance, the dean of a school will have access to all of the information within his/her school and have reports and statistics of its professors, students, publication, etc., but the access to information of other schools may be limited. A professor may want to know about his/her students, publications, project status, and print or distribute curriculum vitae as needed. For external users, they may be prospective students in search for a school, an investor looking for technologies to invest in, a company in search for technologies to license from the university or a ranking or accreditation agency looking for data to calculate their scores. The user may be also a search engine that is launched through the internet to gather data for automated procedures of business intelligence of benchmark exercises. Entities communicate among themselves through the intelligent agents that regulate data flows in order to answer queries or generate reports. This is depicted in figure 6. More details of the knowledge and information system are given in (Cantu et al, 2005).

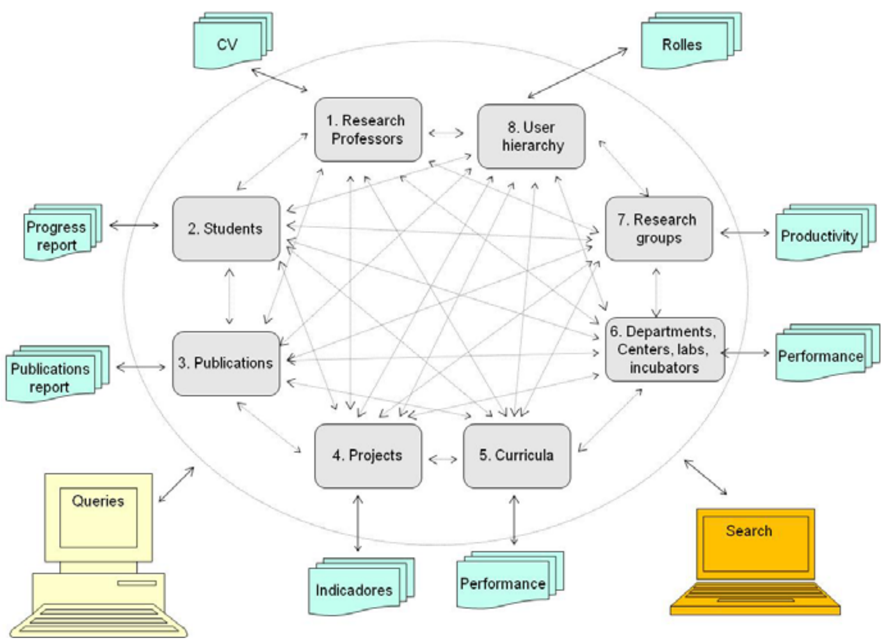


Figure 6. Knowledge and information network repositories

4. THE KNOWLEDGE AND INFORMATION NETWORK PLATFORM

The knowledge and information network approach is supported by a computer platform (KIN platform) that we describe in this section. The KIN platform interacts with the clients described in section 3.3. Among these are the “user” that solicits reports or asks queries about the various entities in the corporate memory and the “admin” user who is responsible for system administration. The KIN platform also interacts with external systems such as web sites or data bases. This is illustrated in figure 7.

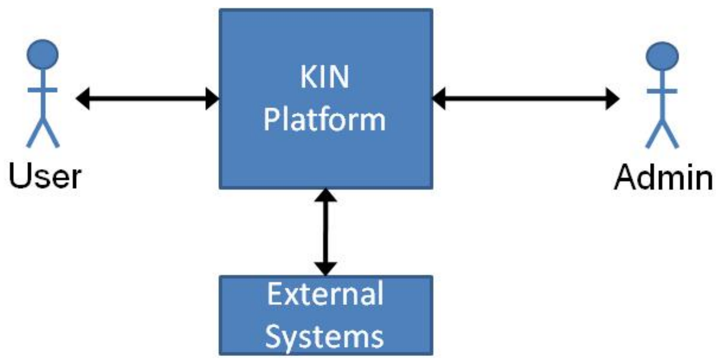


Figure 7. Knowledge and information network platform

The KIN platform is constituted of four main components: an information system, a multiagent system, a knowledge management system and the knowledge to information and information to knowledge interpreter interface. This architecture is shown in figure 8.

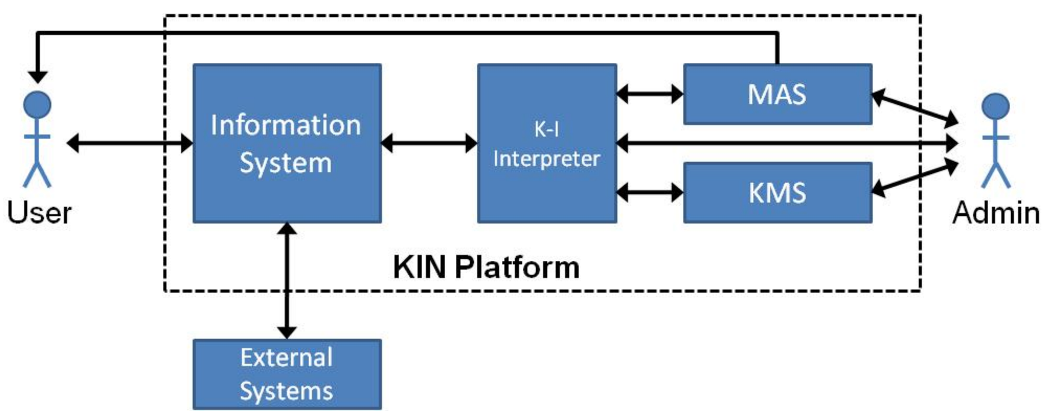


Figure 8. Knowledge and information network platform architecture

4.1 Information System

The information system is responsible for managing the data contained in relational databases. A database management system handles all of the information extracted from other institutional databases and organizes it on the repositories displayed in figure 6. A Model-View-Controller (MVC) framework is used for providing the web interfaces for managing information according to a hierarchical system of permissions. Web services allow for exposing information from the different repositories through a user customized perspective; other institutional web portals or applications consume the information provided by those web services. The current implementation of the information system is built using the PHP programming language, the relational database management system (RDBMS) is attained using MySQL, and Mojavi libraries are utilized for implementing the Model View Controller framework (MVC) (Cantú et al, 2005).

4.2 Multiagent System Architecture

The multiagent system is designed for doing off line operations in the background of the KIN platform (Ceballos and Cantu, 2008).The multiagent system is organized by using three layers of agents: client, repository management and utility. Agents at the client layer administer the interaction with the Admin user and manage responses to users and interact with domain ontologies. Agents at the repository management layer administer the corporate memory by controlling access to the repositories contained in the databases of the information system. Finally, utility agents at the bottom layer provide low-level services to agents in the upper layers. The architecture of the Multiagent System is illustrated in Figure 9.

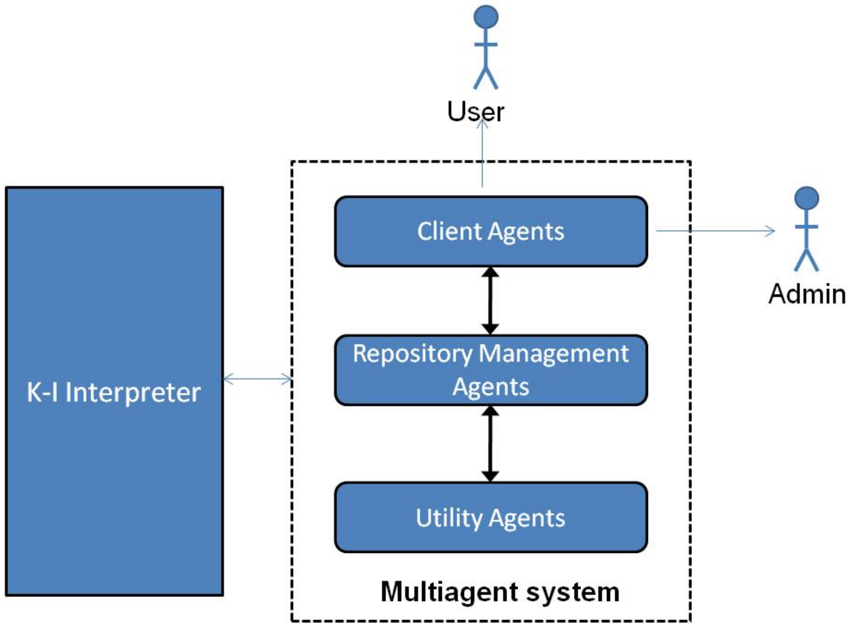


Figure 9. Multiagent system architecture

To perform their tasks, agents do action planning by reasoning with the information contained in the database repositories as well as the domain ontology of the corporate memory which is used to identify objects. They also communicate among themselves and are capable of creating new agents during the planning process as needed. Agents use their plans to do several things: to answer user requests, to provide web services, to perform data mining for extracting useful patterns from databases, to maintain the consistency of the information in the database repositories and to provide services in the background such as information delivery and content auditing.

There are two kinds of agents in the system: *main* agents, who are responsible for providing services through the instantiation and coordination of other agents, and *utility* agents, which perform specialized tasks. A main agent is instantiated by the administrator or by another main agent and is assigned a service specified through a set of goals and objective functions whose satisfaction must be maximized or minimized depending on the task. Information stored in the repository, as well as actions occurring in the information system, is used to generate concrete tasks that the agent delegates to utility agents according to a plan. Should a utility agent be necessary for performing a task and that agent doesn't exist, then, the main agent is capable of creating it by instantiating its type according to a library of agent classes previously defined. The creator-created relationship establishes a hierarchy of agents that speed up the resolution of problems through its decomposition and delegation of concrete subtasks. A utility agent obeys to essential behaviors inherited from its agent class and is capable of committing to achieve goals compatible with its agent class. Its reasoning is based on cases and is encoded in a network of probabilistic cause-effect relationships (Ceballos and Cantu, 2007).

At the Client layer, there are two types of main agents: the user agent and the external agent. User agents interact with the K-I interpreter for converting knowledge into information and vice-versa in responding user queries. They are responsible for modeling user profiles and for communicating with the human user. External users use the Knowledge Management System for mapping concepts from the domain ontology to custom external schemas. The external agent enables an external user for accessing the information through the view built upon such external ontology.

At the repository management layer there is a set of agents for managing each of the database repositories: the guardian agent who maintains the consistency of the information stored in the repository and provides user services by overseeing the work of various agents. For instance: the corrector agent which makes automatic corrections on the information, the publisher agent which expose information of certain repositories through web services interfaces, the auditing agent which conducts system validation and the subscriber agent which alerts users in subscription updates. The agents at this layer are capable of doing Bayesian learning during operation in order to improve their performance in executing their own tasks. Also, there are datamining agents which extract patterns and knowledge from database repositories (Cantu et al, 2006; Rios and Cantu, 2006; Robles et al, 2005; Carrillo et al, 2005)

At the utility layer, there is a set of utility agents which include: the web crawler agent, the web services agent, the email server agent, the database agent, the logs agent, and the ontology manager agent (Ceballos and Brena, 2004; Gottlob and Koch 2004). The agents at each of the three layers of the multiagent system are displayed in figure 10.

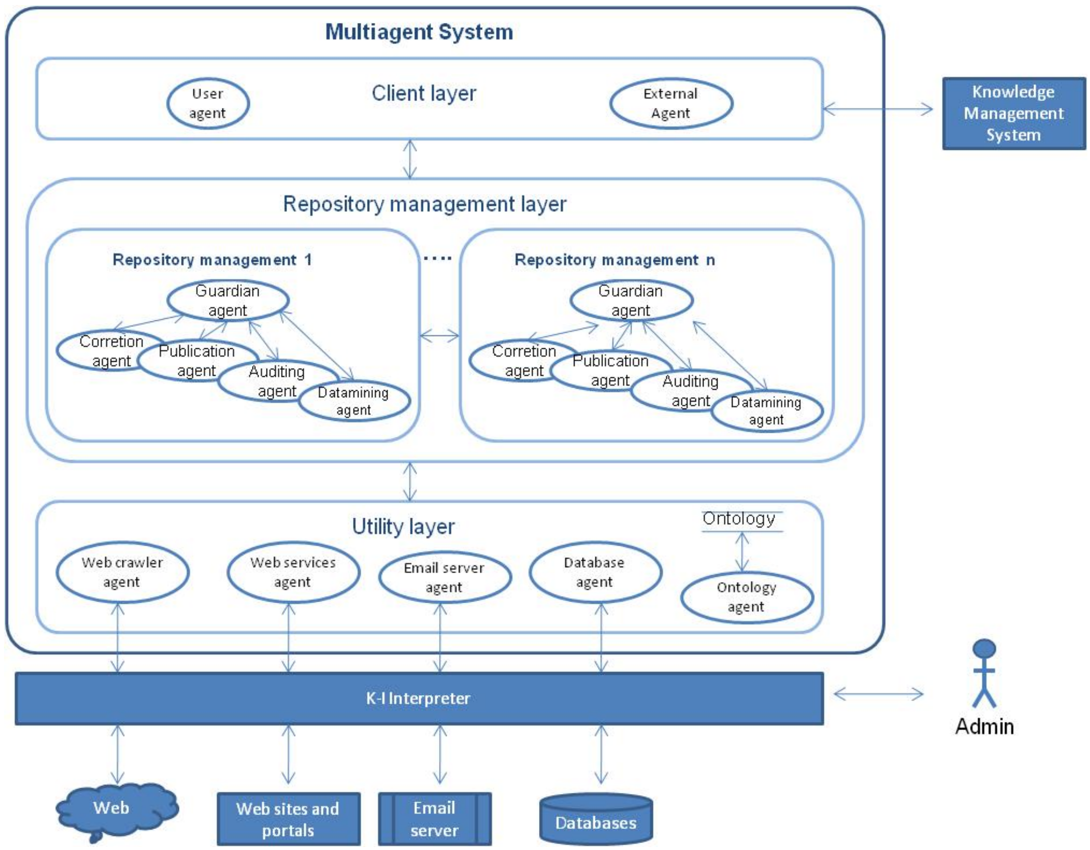


Figure 10. Types of agents at each MAS layer

Interactions among agents on the auditing and correction of the publications repository are shown in figure 11. The log-monitor agent checks periodically the log of the information system and notifies the guardian agent about new publications in the database. The guardian agent instantiates an auditor agent for each auditing rule defined by the expert auditor, and notifies them about the new publication. If an inconsistency is detected the auditor agent chooses between notifying a corrector agent or notifying the expert or the corresponding author through their respective user agent. Notified by email, users can accept or reject the correction or make a direct change of the record through the repository's web interfaces provided by the information system.

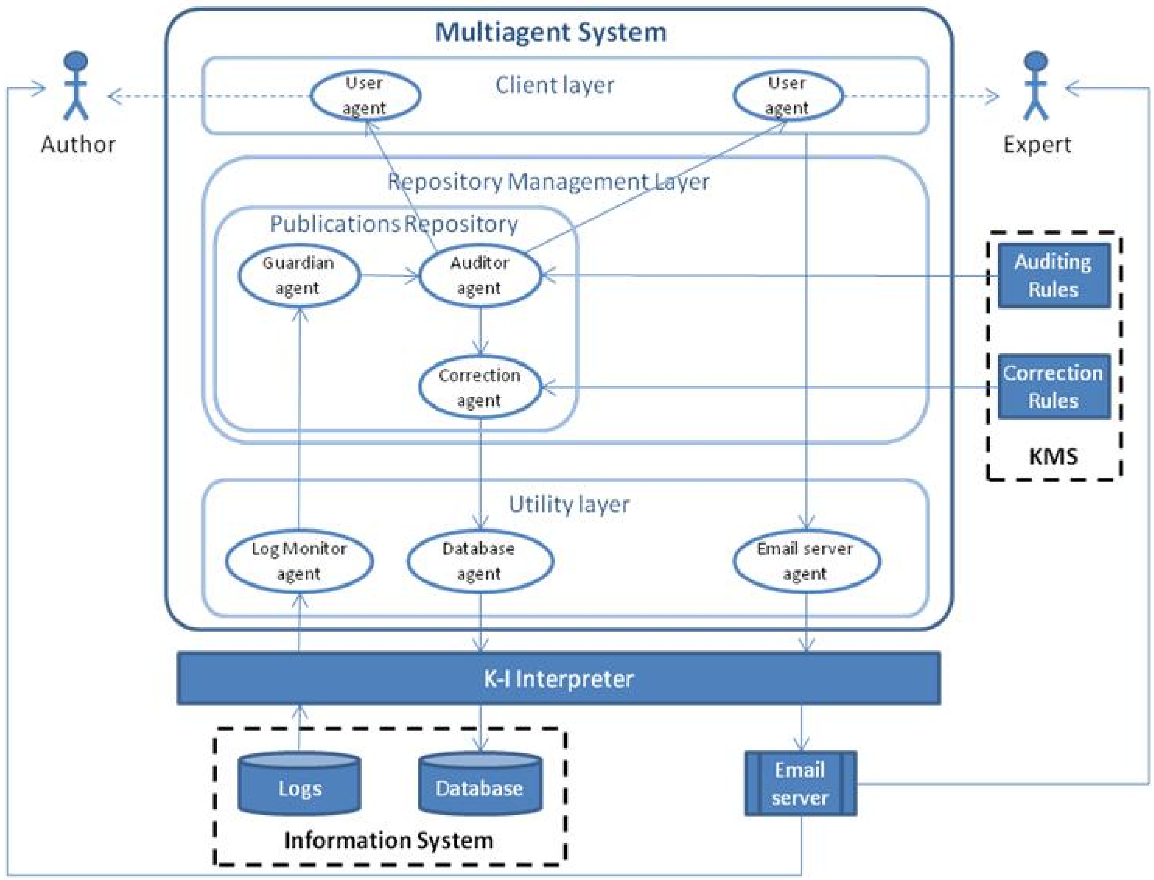


Figure 11. Multiagent system scenario for publications auditing

The Web Ontology language OWL[7] is used for representing classes of agents, properties and actions the agent can perform. The consequences of the actions are described by rules that are used for configuring plans consistent with agents' goals. The multiagent system architecture is implemented using the JADE platform.

4.3 Knowledge Management System

The Knowledge Management System (KMS) is responsible for managing the knowledge tasks of the KIN platform. Among the knowledge tasks are the administration of the domain ontology, knowledge representation mechanisms and other knowledge management tasks. Knowledge administration is attained through various mechanisms which include a domain ontology manager, information-knowledge mappings management, a data-source manager,

[7] http://www.w3.org/TR/owl-features/

among others. The KMS is constituted by a set of ontological repositories and a tool for their management. The Protégé Ontology Editor and Knowledge Acquisition System[8] is used for managing the ontological repositories. Ontological repositories are codified through the OWL language and the resource description framework RDF[9]. The KMS manages five ontological repositories: the research corporate memory ontology (RCMO), the public research corporate memory ontology (P-RCMO), internal and external data sources, information-knowledge mappings, and external ontologies. RCMO describes the elements of the system classified according to the taxonomies shown in Figures 3-5, constituting the domain ontology. P-RCMO contains concepts and relationships that can be accessed by external users. Concepts defined on P-RCMO are mapped to concepts of the original RCMO. Data sources are represented by individuals describing the connection to relational databases or websites. Information-knowledge mappings, such as the "database to resource description framework" mapping (D2R) (Bizer, 2003), allow for a transformation between information stored in the registered data sources and the RCMO ontology. Finally, external ontologies are defined by external users for extracting information from the KIN repositories through mappings to the Public RCMO. This architecture is shown in figure 12.

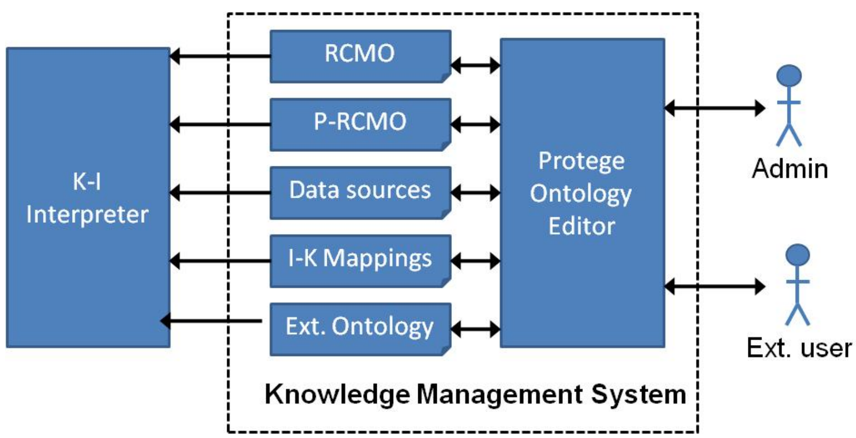


Figure 12. The Knowledge Management System

Auditing and correction rules are stored at the RCMO ontology. Auditing rules validate: 1) duplicated records, and 2) the internal consistency of records. Correction rules relates the

[8] http://protege.stanford.edu/

[9] http://www.w3.org/RDF

inconsistency with a correction action like replacing or setting some attribute of the record or deleting the record itself from the database. On the same way, is stored each evaluation of auditing and correction rules allowing to determine probabilistically how effective are on a given case. Cases are codified using attributes contained on publications metadata.

4.4 Knowledge-Information Interpreter

The Knowledge-Information interpreter is responsible for mapping information into knowledge and vice-versa. The data and information from relational databases is mapped into schemas represented by semantic web ontologies in order to provide meaning to syntactic terms used by the multiagent system. Mappings are updated through the KMS as described in section 4.3. A database to resource description framework translator such as the D2R Server[10] uses such mapping to enable utility agents accessing data source repositories through user queries. On the opposite way, utility agents use parameterized commands from a library for updating, deleting and inserting information on the data sources. Also, external data and information from external sources are mapped into terms of the RCMO ontology. This is shown in figure 13.

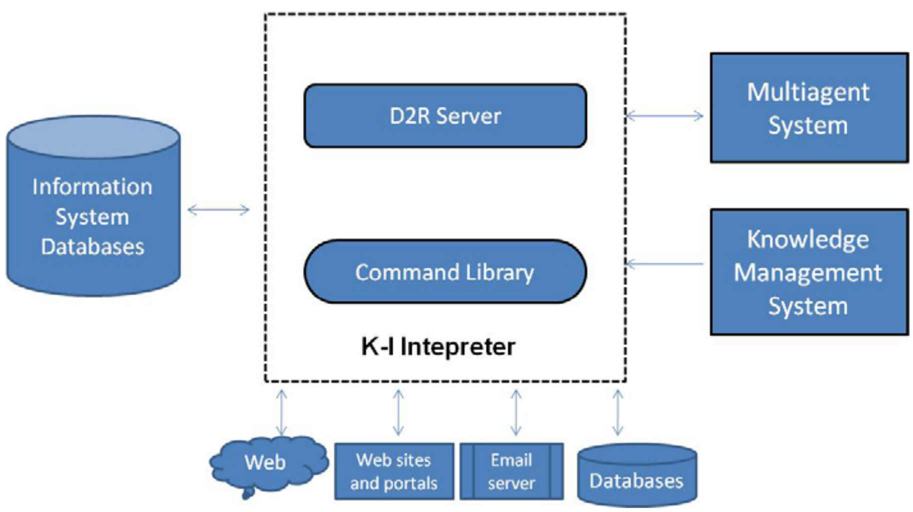


Figure 13. The Knowledge-Information Interpreter

[10] http://www4.wiwiss.fu-berlin.de/bizer/d2r-server/

4.5 Bayesian Causal Learning

We describe a kind of agent that is capable of learning parameters of rules by using Bayesian causal networks (Pearl, 2000). The correction agent and the auditing agent that are used for keeping consistency within the knowledge repositories are of this type. Initially, consistency rules are supplied by Admin user and evolve during system operation by Bayesian learning. For example, publications registered by researchers are audited and complemented by agents. Users are notified the result of auditing by an agent email and they can accept, modify or reject such changes. Inconsistency detection is based on stored rules that describe inconsistency patterns. Figure 13 shows a causal network of the semi-supervised process for auditing with three possible outcomes. First, if there is correction available, the correction is done on-line automatically by applying an available rule. Second, if the inconsistency is not identified or if the confidence of the correction does not reach a threshold then a human expert can be consulted to apply the correction. Third, if the inconsistency cannot be solved automatically or by the expert, the user responsible for the publication is notified about a potential inconsistency. For the first two cases, the user is notified in order to give him/her the opportunity to revise the correction. In any case, user agents monitor actions taken by the user with respect to publication inconsistency. Finally, the confidence on the publication is calculated in terms of the inconsistencies detected, the corrections made and the reinforcements of experts and users. Audited cases are stored and used for making parameter learning. On this way the system modifies its decisions according to its experience (Ceballos and Cantu, 2008).

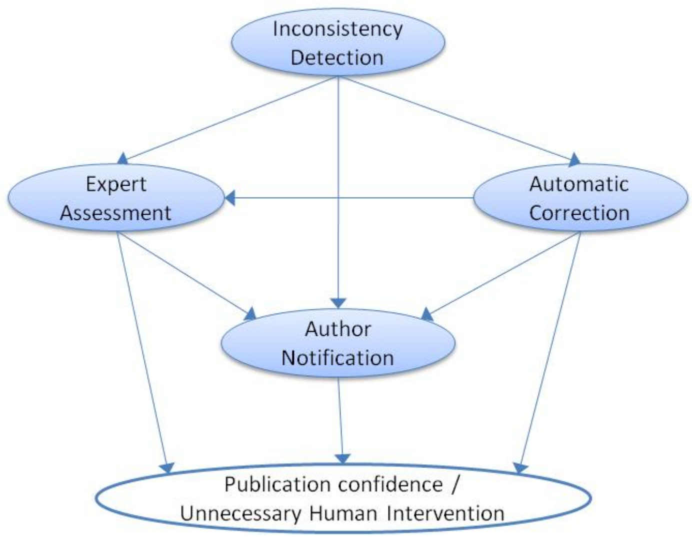


Figure 13. Causal network of automatic repository consistency maintenance

In this way, the multiagent system contributes to the goal of maximizing the confidence on the repository by keeping the knowledge repositories consistent minimizing human intervention.

Data mining facilities have also been added to extract knowledge from the database repositories. For instance

4.6 System implementation

The information system described in section 4.1 was developed during the year 2003 and has been fully operational since August 2004. The multiagent system architecture, the knowledge management system, the Knowledge-Information interpreter and Bayesian causal learning describes in sections 4.2-4.5 were developed in 2006 and are operational since 2009. More specifically, auditing and correction rules were modeled using Alloy (Jackson, 2006) and CCalc (Giunchiglia et al, 2004) as described in (Ceballos and Brena, 2008). The auditing process was modeled using the Electronic Institution formalism (Sierra and Noriega, 1997) (Esteva et al, 2002) for which an experimental prototype was implemented (Ceballos et al, 2009).

5. CASE STUDY

In this section we describe a case study using Tecnológico de Monterrey (ITESM) as an example of a knowledge organization by using an action research approach (Reason and Bradbury, 2004). ITESM is a university with operations in 33 campuses in México, sixty thousand undergraduate students, twelve thousand graduate students, eight thousand faculty members, of which around twelve hundred are research professors. In 2002 an institutional research program was initiated that would change ITESM profile from a teaching oriented university into a teaching a research university. To achieve this goal, financial resources were allocated to foster research in the various schools of its main campus. Research groups were established and supported with seed funds to support professors to lower teaching loads, to pay students tuition fees and research assistantships, traveling expenses, materials and research infrastructure. The areas of research were defined from market studies and the expertise of human resources to include biotechnology, health, mechatronics, nanotechnology, information technologies, telecommunications and resources such a energy, water, air, forests, etc. A research group is constituted of about 18 researchers including the principal investigator, adjunct professors, postdoctoral researchers, and doctoral, master and undergraduate researchers. As a result, research products proliferated, research students became alumni and the doctoral programs were revitalized (Cantu et al, 2009). Soon, the need for managing the assets generated from research activities became evident. Policies and regulations were established and knowledge management procedures were adopted to manage the research assets including a computer platform and a corporate memory that would assist officers and researchers in decision making. The concept of multiagent-based knowledge networks that had been developed at ITESM was revised and extended as a knowledge and information network to be used in the management of research assets (Aguirre et al, 2001).

5.1 Knowledge and information network at ITESM

The network and information network concept described in section 3 emerged from the analysis of the situation at ITESM in which external and internal sources were identified as shown in figure 1. The network elements were classified as shown in figure 2: knowledge products, human resources and research programs.

For the knowledge products the main tool was the system to manage publications. The system was built around the concept of a *knowledge pump* corporate memory which gathers the research products generated by professors and students at the various research groups and academic programs at the university. The taxonomy displayed in figure 3 was designed and used to classify the types of publications. This system proved useful in establishing the scientific merit of publications as well as the potential for economic profits derived from its transferring as in the case of a patent or any kind of invention or the initiation of a start-up company. Currently, the publications corporate memory stores around 38,000 research products of the types shown in the taxonomy that have been generated by ITESM researchers since year 2000.

For the human resources element the main tool was the system to manage researchers of type defined in the taxonomy shown in figure 4: professors, associates, assistants, etc. Each type is identifies by a set of attributes that the system organizes and from which various reports can be generated. For instance, the system is able to generate an up to date curriculum vitae of each researcher in the data base including his or her publications and research achievements. Currently, the human resources corporate memory stores around 13,200 researchers of the types shown in the taxonomy including 2,700 research professors, 3,740 research students and 6,740 external authors, that have generated some kind of publication at ITESM since year 2000.

For the research programs element the main tool was the system to manage the units of the taxonomy displayed in figure 5: curricula, departments, projects, etc. Currently, the research programs corporate memory stores around 3200 records of the types shown in the taxonomy including 70 research centers, 120 research chairs and around 3000 research projects performed at ITESM since year 2000.

The knowledge and information elements interact among themselves as shown in figure 6. For instance, the information system relates those elements in the following way:

- The system receives information about university human resources by means of an automated on-line interface to the Banner Enterprise Resource Planner (ERP) system that contains payroll, registry and accounting databases.
- The system inputs information from researchers through a semi-automated user interface with pull-down menus. For instance, if the professor feeds in a journal article, the information attributes of that journal are automatically fetched from the journals database in the system.
- This database contains the main journals of the scientific disciplines as well as information about its impact factor and other attributes. The journal database is automatically updated when a new article is published in a journal that is not yet in it. Every new publication is audited by a human expert to make sure it is correctly classified and all the relevant information has been fed in.

The distribution of information and knowledge is an active process done as follows:

- The system generates automatically predefined reports of publications of various types as well as statistics and indicators by professor, student, department, college, school, or campus. One of these reports includes the curriculum vitae (CV) of any researcher. It also has the capacity for updating personal and departmental web pages through web services.
- The system sends automatic notifications to a co-author of a paper when the author feeds in the paper into the system. It also sends professors automatic messages when a publication that is relevant to his/her research interests is uploaded.

User interfaces for capturing assets were designed for every type of product and contain different attributes depending on the nature of the product. They appear on the principal menu in order of importance; for example, *Journal Article* appears in first place, whereas *Program Committee Member* appears as part of *Support Activities*. For example, author association to a publication is made through a custom interface. Figure 14 shows the author association interface displayed over the Journal Article register screen.

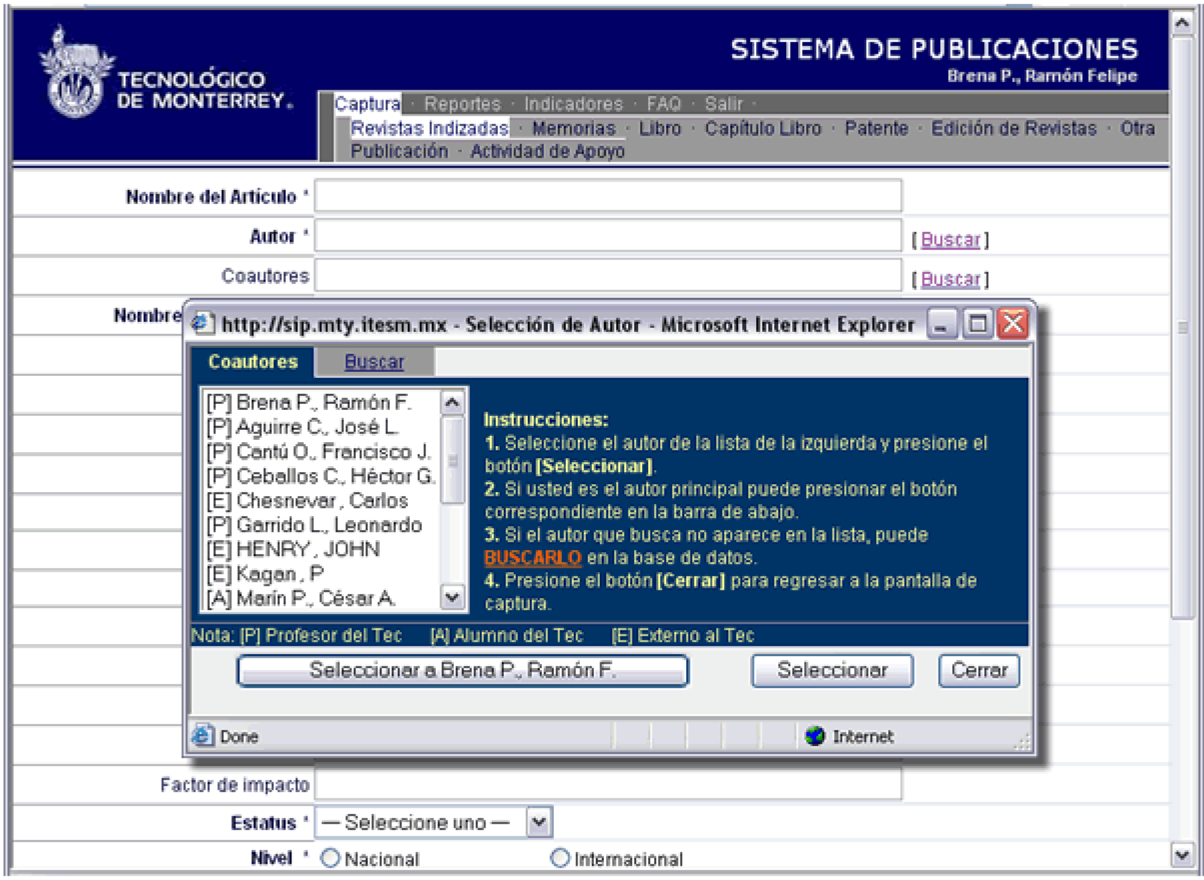
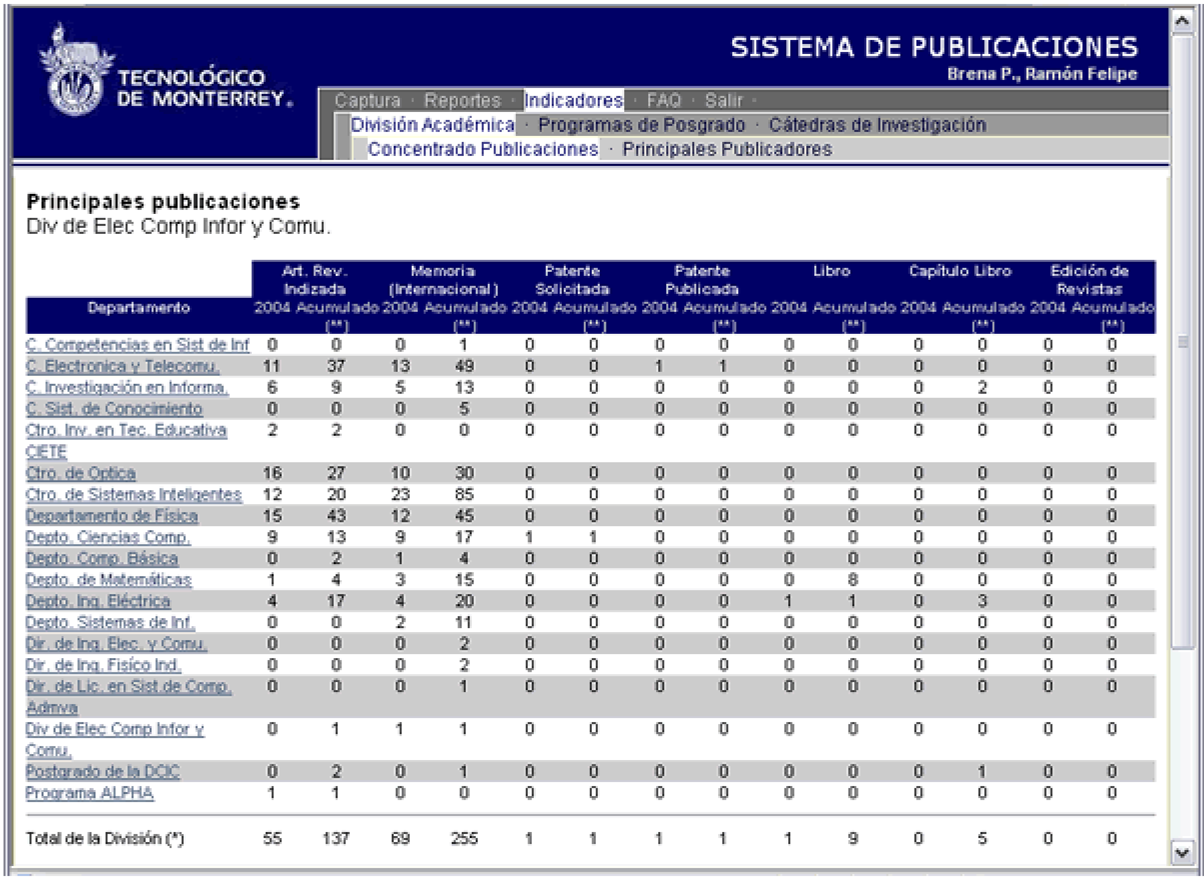


Figure 14. Author association interface

Statistical reports are generated in a more elaborated way that listing reports. Two main considerations in their design were (1) generating reports according to the taxonomy categories and (2) the ability to drill down and roll up the information. First, dimensions were defined according to the hierarchy of concepts over which it is possible to navigate; for example, in an organizational report the dimension is given by Campus => School => Department => Person. Then SQL templates were defined in which scope, publication type and detail level are passed as parameters to a function that builds the query. Results are given in a standard format letting build matrixes that conform the indicator report. There is a PHP template that formats the results and creates links for navigation throughout reports. An example of an indicator report is shown in Figure 15.



Principales publicaciones
Div de Elec Comp Infor y Comu.

Departamento	Art. Rev. Indizada 2004	Art. Rev. Indizada Acumulado (**)	Memoria (Internacional) 2004	Memoria (Internacional) Acumulado (**)	Patente Solicitada 2004	Patente Solicitada Acumulado (**)	Patente Publicada 2004	Patente Publicada Acumulado (**)	Libro 2004	Libro Acumulado (**)	Capítulo Libro 2004	Capítulo Libro Acumulado (**)	Edición de Revistas 2004	Edición de Revistas Acumulado (**)
C. Competencias en Sist de Inf	0	0	0	1	0	0	0	0	0	0	0	0	0	0
C. Electronica y Telecomu.	11	37	13	49	0	0	1	1	0	0	0	0	0	0
C. Investigación en Informa.	6	9	5	13	0	0	0	0	0	0	0	2	0	0
C. Sist. de Conocimiento	0	0	0	5	0	0	0	0	0	0	0	0	0	0
Ctro. Inv. en Tec. Educativa CIETE	2	2	0	0	0	0	0	0	0	0	0	0	0	0
Ctro. de Optica	16	27	10	30	0	0	0	0	0	0	0	0	0	0
Ctro. de Sistemas Inteligentes	12	20	23	85	0	0	0	0	0	0	0	0	0	0
Departamento de Física	15	43	12	45	0	0	0	0	0	0	0	0	0	0
Depto. Ciencias Comp.	9	13	9	17	1	1	0	0	0	0	0	0	0	0
Depto. Comp. Básica	0	2	1	4	0	0	0	0	0	0	0	0	0	0
Depto. de Matemáticas	1	4	3	15	0	0	0	0	0	8	0	0	0	0
Depto. Ing. Eléctrica	4	17	4	20	0	0	0	0	1	1	0	3	0	0
Depto. Sistemas de Inf.	0	0	2	11	0	0	0	0	0	0	0	0	0	0
Dir. de Ing. Elec. y Comu.	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Dir. de Ing. Fisíco Ind.	0	0	0	2	0	0	0	0	0	0	0	0	0	0
Dir. de Lic. en Sist.de Comp. Admva	0	0	0	1	0	0	0	0	0	0	0	0	0	0
Div de Elec Comp Infor y Comu.	0	1	1	1	0	0	0	0	0	0	0	0	0	0
Postgrado de la DCIC	0	2	0	1	0	0	0	0	0	0	0	1	0	0
Programa ALPHA	1	1	0	0	0	0	0	0	0	0	0	0	0	0
Total de la División (*)	55	137	69	255	1	1	1	1	1	9	0	5	0	0

Figure 15. Statistical report

5.2 Results

The Knowledge and Information Network system just described has been in operation since August 2004 and has proved a useful tool in assisting university officers and researchers in having ITESM move towards a teaching and research university and generating value from its knowledge assets. Table 1 shows key parameters of this transition since 2002 to 2009 and sets goals for year 2015.

Table 1. ITESM Research statistics

Attribute	**2002**	**2009**	**2015**
Research groups	20	125	150
Accredited researchers	100	270	500
SCI journal articles	117	350	500
Books	25	109	200
Patents	1	50	100
Technology Licensing	1	20	100
PhD students	100	520	1200
Technology-based spinoffs	1	30	100
TIMES Higher Education-QS ranking	433	338	200
Research funding USD million	13	57	100

Research groups have grown from 20 to 125 since 2002 and are expected to be around 150 by 2015. Researchers may obtain a national accreditation from the National Council for Science and Technology en Mexico (CONACYT) based on scientific merits. Journal articles in the Science Citation Index and the Social Science Citation Index have tripled in this period and the number is expected keep increasing in the following five years as well as books authored by faculty and published by editorial firms. Patents filed per year were almost non-existent in 2002 whereas the number is now around 50 in 2009 and may double by 2015 while the licensing of those patents through the Office of Technology Transfer is also taking off. The number of PhD students which is a key element of any university’s research strategy, has multiplied times five and will double in the following years. The number of technology-based spinoff companies incubated within the research groups has had a considerable increase and will continue growing in the near future. The TIMES Higher Education World University Ranking has moved from position 433 in 2007 to position 338 in 2009 and the goal is to be in the top 200 by

2015. Finally, all research activities have been supported by funding from internal and external sources, the latter including research grants and industrial research contracts.

6. CONCLUSIONS

We described a knowledge and information network approach to handing research assets at a knowledge organization. We believe that institutions which adopt such an approach obtain multiple benefits from it; for instance, documenting the research products in a corporate memory, gaining awareness about the research assets, and keeping track of their scientific impact as well as of the economic and social benefits that may be obtained from them. The knowledge and information network system for managing research products at ITESM has proved useful in gathering research products as well as in the distribution of the knowledge generated by the research activities. Researchers are now conscious about what are the important journals and conferences in their disciplines and are teaching and encouraging their students to write their theses and projects results in scientific papers that are peer-reviewed and criticized during conference oral presentations. Professors from technological disciplines are also filing patents from their innovations and teaching their students to do so. Research centers and departments have access to historical reports that show statistics about their scientific productivity and main weaknesses. Graduate programs and schools are using the indicators and reports generated by the system to obtain accreditations from scientific societies, accreditation boards as well as governmental and international funding agencies. Students and professors are starting spin-off companies that commercialize products designed as a result of their research and innovation activities utilizing the support and infrastructure provided by the entrepreneurial program at ITESM. At the university level, there is useful information provided by the system that indicate what are the research areas with more publications, who are the researchers with the highest productivity, what are the journals and conferences in which researchers publish the most, what is the proportion of papers written jointly by professors and students, what theses produce published papers, what publications are obtained from research grants, what graduate programs are generating scientific papers, and similar questions. Another benefic of *the system* is the awareness that external entities will have about the university research and innovation capabilities. Among these entities are companies and corporations that demand research services, funding agencies and prospectus students that look for challenging research projects and research environments. Benefits of such a system are particularly important because promotes the emergence of technology-based spinoffs and the development of knowledge-based economies.

We have followed an action research methodology for reflecting about the processes and products generated by research activities as well as the economic value associated to such knowledge assets. We have followed a knowledge management methodology and developed a knowledge pump corporate memory for storage and distribution of research knowledge assets through a multiagent-based computer information system. Being operational since August, 2004, this system has already proved its usefulness in creating an awareness and consciousness about the importance of research assets and measuring its relevance in terms of both its scientific impact and the creation of wealth.

Related work

The use of a system like this is not different from what is done in other parts of the world since the countries face similar problems due to the globalization phenomena. However, universities that follow approaches similar to the ones described in this paper are better off in contributing to their local economies. Many companies and government agencies have developed corporate memories and knowledge management systems to handle their knowledge assets (Liebowitz and Beckman, 1998; Liebowitz, 2004). Among those companies are General Electric, Monsanto, Buckman Laboratories, Skandia, Blue Cross, Federal Express, KPMG and many others. On the other hand, there are various companies like Thompson Scientific, Google, Elsevier, Springer-Verlag and other publishers that sell information in data bases of scientific publications as their core business. However, to the best of our knowledge, the system described in this paper is one of the first integrated solutions developed for a university environment that is currently operational.

ACKNOWLEDGEMENTS

The authors acknowledge Tecnológico de Monterrey at Monterrey and its President Dr. Alberto Bustani for the support and academic freedom for doing research that lead to the knowledge and information network concept, the multiagent system and its implementation. We also acknowledge the members of the research and graduate programs, and the staff of the information technologies offices for their support in the system design and implementation. We acknowledge the deans and chairs of the various schools, departments and research centers as well as the faculty members and graduate students on Campus for their assistance in deploying the system and making it operational. Finally, we acknowledge the very valuable suggestions of anonymous referees that contributed to significantly improve and shape this paper.

REFERENCES

1. Aguirre, J.L., Brena, R. Cantu-Ortiz, F. (2001). Multiagent-based Knowledge Networks. *Expert Systems with Applications an International Journal*, No. 20, pp 65-75, Elsevier (ISSN 0957-4174)
2. Alvarado, M., Sheremetov, L., Bañares-Alcantara, R., & Cantú-Ortiz, F. J. (2007). Current challenges and trends in intelligent computing and knowledge management in industry. Knowledge And Information Systems, 12, 117-127. UK
3. Bizer, Christian (2003). D2R MAP – A Database to RDF Mapping Language. In proceedings of The Twelfth International World Wide Web Conference (WWW2003).
4. Cantú F., Ceballos H., Mora S.P. and Escoffié M. (2005). A Knowledge-based information system for managing research programs and value creation in a university environment. Americas Conference on Information Systems – AMCIS. USA. pp.781-791.
5. Cantú, F., Garza, L., Robles A. and Morales R. (2006). Learning and using Bayesian networks for diagnosis and user profiling. Twelve Americas Conference on Information Systems - AMCIS . Acapulco, México
6. Cantú, F., Bustani, A, Molina A., Moreira H. (2009). *A knowledge-based Development Model: The Research Chairs Strategy*. Journal of Knowledge Management. Reino Unido. pp: 154-170.
7. Carrillo, M.A., Cantu, F.J., Garza, L.E., Morales, R. (2005). Learning Bayesian Network Structures from Small Datasets using Simulated Annealing and Bayesian Score. IASTED Conference on Artificial Intelligence Applications, Austria
8. Ceballos H. and Brena R. (2004). Finding Compromises between Local and Global Ontology Querying in Multiagent Systems. Lecture Notes in Computer Science (LNCS). Germany, pp: 999-1011.
9. Ceballos H. and Brena R. (2008). Modeling Domains and Multiagents Systems with Alloy and CCalc: a Case Study on Publications Auditing. CCIR Technical Report. Tecnologico de Monterrey. México. December 2008.
10. Ceballos, H. and Cantú F. (2007). Integrating Semantic Annotations in Bayesian Causal Models. 20 International Workshop on Description Logics (DL'07). Bozen Bolzano University Press. pp: 527-528.

11. Ceballos, H. and Cantú F. (2008). Modeling intelligent agents through causality theory. Proceedings of the Sixth Mexican International Conference on Artificial Intelligence. IEEE Computer Society Press. pp: 201-210.
12. Ceballos H., Noriega P. and Cantú F. (2009). Autonomic Information Auditing through Electronic Institutions. IIIA-CSIC. Technical Report IIIA-TR-2009-07. Spain. July 2009.
13. Esteva M., de la Cruz D., and Sierra C. (2002). ISLANDER: an electronic institutions editor. In First International Conference on Autonomous Agents and Multiagent systems, pages 1045-1052, Bologna, July 2002. ACM Press.
14. Etzkowitz, H. (2001). The second academic revolution and the rise of the entrepreneurial science. IEEE Technology and Society, Vol. 20, No. 2.
15. Garfield, E. (1999). Journal impact factor: a brief review. Canadian Medical Association Journal. 161(8), p.979-80.
16. Giunchiglia E., Lee J., Lifschitz V., McCain N. and Turner H. (2004). Nonmonotonic Causal Theories. Artificial Intelligence, Vol. 153, pp. 49-104.
17. Gottlob G, Koch C (2004). Monadic datalog and the expressive power of languages for web information extraction. *JOURNAL OF THE ACM*. 51 (1), USA, ACM Press, 74-113
18. Gutierrez, F. J., Lerma, M., Salgado, L. R., Cantu, F. J. (2002). Biometrics and Data Mining: Comparison of Data Mining-Based Keystroke Dynamics Methods for Identity Verification, MICAI-2002, Lecture Notes in Artificial Intelligence 2313, Springer-Verlag
19. Jackson D. (2006). Software Abstractions: Logic, Language and Analysis. MIT Press.
20. Liebowitz, J.and Beckman, T. (1998). Knowledge Organizations. Saint Lucie Press Washington, DC
21. Liebowitz J. (1999). Knowledge management Handbook. CRC Press LLC, Boca Raton FL
22. Liebowitz J. (2004). Addressing the Human Capital Crisis in the Federal Government: A Knowledge Management Perspective. Burlington MA: Elsevier
23. O´Leary, J., Quacquarelli, N. Ince, M. (2009). Top Universities Guide. Quacquarelly Symonds Ltd, London, UK
24. Padgham L. and Winikoff M. (2004). The Prometheus methodology. Chapter in Methodologies and Software Engineering for Agent Systems. Springer. Vol. 11.
25. Pearl J. (2000). Causality. Models, Reasoning, and Inference. Cambridge University Press.
26. Reason, P. and Bradbury, H. (ed.) (2004). Handbook of Action Research: Participative Inquiry and Practice, SAGE Publications, LTD, London, UK
27. Rios María Teresa and Cantú Ortiz Francisco J. (2006). Knowledge Discovery in Academic Registrar Data Bases using Source Mining: Data and Text. Twelve *Americas Conference on Information Systems - AMCIS* . Acapulco, México.
28. Robles, A., Cantu, F.J., Morales, R. (2005). A Bayesian Reasoning Framework for On-line Business Information Systems. Proceedings of the Eleventh Americas Conference on Information Systems, Omaha, NE, USA August 11th-14th.
29. Sierra C. and Noriega P. (1997). A formal framework for accountable agent interactions. In Fifth Bar-Ilan Sympossium on Foundations of Artificial Intelligence, pages 23-24, Ramat-Gan, Israel, June 1997.
30. Uschold, Mike and Gruninger, Michael (1996). Ontologies: Principles, Methods and Applications. Knowledge Engineering Review, 11(2)

31. Weiss G. (1999). Multiagent Systems: A modern approach to distributed artificial intelligence. The MIT Press. England.
32. Wooldridge M., Jennings N. and Kinny D. (2000). The Gaia Methodology For Agent-Oriented Analysis And Design. Journal of Autonomous Agents and Multi-Agent Systems. 3, pp 285-312.
33. Zuckerman, M. (2004). America's Best Graduate Schools. U.S. News & World Report. USA