Expert Systems with Applications 37 (2010) 5272–5284

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

A multiagent knowledge and information network approach for managing research assets

Francisco J. Cantú *, Héctor G. Ceballos

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

A R T I C L E I N F O

Keywords:
Knowledge-based systems
Multiagent systems
Knowledge management
Research assets
Ontology
Entrepreneurial science

A B S T R A C T

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 developed by the research groups in a knowledge institution. The 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-based books, patents, technology licensing, trademarks, incubation of technology-based startup companies and others. Intellectual capital consists of the talent and expertise of research staff, such as professors, students and researchers. Research programs include academic curricula, research units, research infrastructure, and business incubators. The approach is supported by an intelligent platform that contains an information system, a multiagent-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 for distributing existing research assets both within the organization and beyond, a variety of research reports, online consultations, a search engine, web services and data mining facilities for knowledge extraction. The knowledge and information generated by the system guide managers in defining strategies on competitiveness, for example, 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 knowledge of research capabilities and for stimulating entrepreneurial science initiatives.

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's welfare are measured in terms of the level of investment in science and technology allotted by their governments. Generating knowledge is a necessity in modern economies and, subsequently, the most important aspect is the economic value that can be attained from the knowledge assets created by key players in knowledge economies. Economic value is often measured by the number of technology-based companies, the number of jobs requiring 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 21st century play an important role as they become engines of economic development for the society (Etzkowitz, 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 in universities, companies and government-supported laboratories. This knowledge is socially validated by the scientific community after a peer-review process and made public by publication in journal articles, conference papers, technical reports, magazine articles and other means. In contrast, inventions and innovations are kept secret until intellectual protection is obtained through a patent, a trade-mark, a distinctive sign, industrial secrets or other intellectual property mechanisms. The scientific productivity of an organization as well as that of the organization's researchers is commonly measured in terms of the quality of the journals in which they publish. Quality is determined by the journal's indexes and impact factor, which depends on the number of citations in articles in other journals. For inventions and innovations, the measure is economic impact, which derives from licensing, spin offs and other financial measures. Combining both types of knowledge appears to be a

* Corresponding author. Tel.: +52 81 8328 4182; fax: +52 81 8158 2264.
E-mail addresses: fcantu@itesm.mx (F.J. Cantú), ceballos@itesm.mx (H.G. Ceballos).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.01.012

challenge for knowledge-based organizations. In our research, we focus on the management of scientific and innovation knowledge in organizations such as universities and technology-based companies (Alvarado, Sheremetov, Bañares-Alcantara, & Cantú-Ortiz, 2007). The main thrust is the production of benefits from the knowledge assets generated by scientific research and development.

Integrating current knowledge and information technologies to manage knowledge in novel ways in modern organizations is a key aspect of this article, which is organized as follows: Section 2 gives 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 offers the conclusions and future work.

2. Background and related work

Knowledge-based organizations build their competitiveness on the market share of the products and services they offer and from the value perceived in them by their potential customers. This value often comes from the organization's technology research and innovation areas, which are composed of design and engineering teams for implementation and followed by a marketing plan for launching new or improved products or services on the market. This is also true for organizations such as colleges and universities, which emphasize research and development as part of their mission statement. For instance, universities and graduate schools worldwide are annually ranked by various agencies using various indicators. These rankings often have a media impact and are used to promote student enrollment as well as for obtaining research grants. Research evaluation is an important component of rankings calculated from weighted indicators that yield a general score. Typical indicators include the quality of journal articles and number of citations, faculty awards, the educational background of graduate students and researchers, research expenditure and, nowadays, the number of startup companies arising from the research and development work of the faculty and students. Research evaluation has created the necessity for information systems to assist administrators in integrating data, information and knowledge in 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 measuring the quality of a 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 winners of Nobel Prizes and 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 (ÓLeary, Quacquarelli, & Ince, 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 in 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] provides this service for free, obtaining the information through automatic web data extraction. Patent information is 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 quantity 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 surrounding knowledge, such as its creation, storage, distribution and use (Liebowitz, 1999). In our approach, all this information is managed by a knowledge-based organization using KM methodologies like the one proposed by Liebowitz (Liebowitz & Beckman, 1998). The KM methodology is composed of 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 that stores and distributes information and knowledge relevant for business operation. Corporate memories are classified according to the *capture* and *distribution* mechanisms 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 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 considered intelligent if they pursue their goals and optimize their performance according to 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 a part achieves a set of global objectives through the interaction of the various types of agents. An arrangement of agents is called a Multiagent System (MAS) (Weiss, 1999).

There are several methodologies for developing MAS; one of them, GAIA (Wooldridge, Jennings, & Kinny, 2000), based on KM methodologies, specializes in representing a human expert's knowledge, and another, Prometheus (Padgham & Winikoff, 2004), 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 kinds of agents, software and human (Sierra & Noriega, 1997).

Description Logics (DL) is 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 1980s for representing knowledge. Ontology is a kind of dictionary that defines the semantics of terms in a problem domain and is used by agents to reason about domain situations (Uschold & 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 the reverse process is an important component of knowledge-based systems. Information obtained from data by data processing mechanisms is well understood in business information systems, whereas

[1] http://ed.sjtu.edu.cn/ranking.htm.
[2] http://www.thomson.com/.
[3] http://scholar.google.com/.
[4] http://www.uspto.gov/.

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 framework (RDF) for accessing vast amounts of data in web sites and databases.[6]

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

3. A 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 human 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 is maintained for handling the complexity of the interactions among entities at various levels of design and implementation. Repositories are found at both extranet and intranet levels. The extranet level repositories include public databases and document and scientific publication web sites 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. Fig. 1 depicts the KIN and its context.

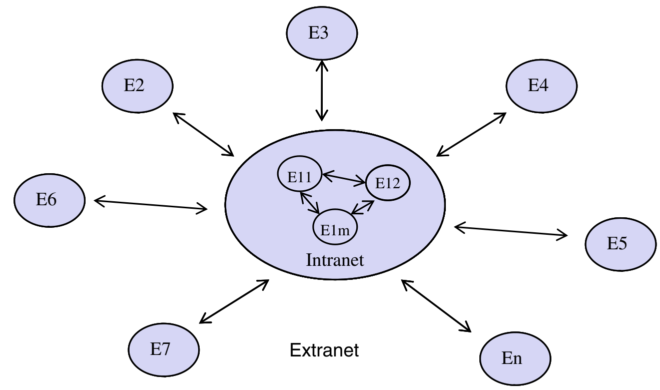

Fig. 1. Knowledge and information network with internal and external elements.

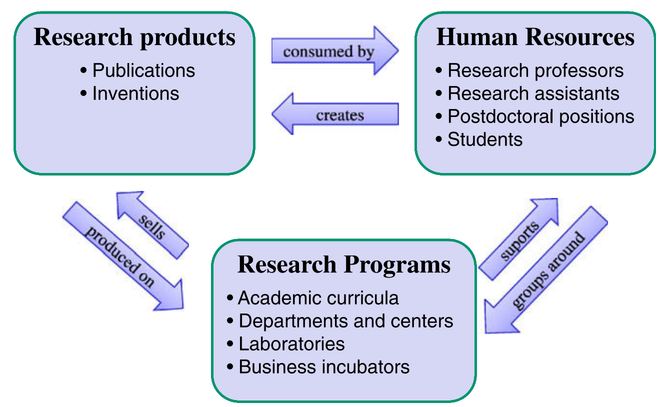

Fig. 2. Conceptual model of the knowledge and information network.

3.1. Network components

The knowledge and information network contains three types of components: research products, human resources or intellectual capital, and 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 are mainly researchers – faculty members, research assistants, postdoctoral participants and graduate students, as well as technical and administrative staff members. The research programs are of different types; for a university they contemplate graduate curricula and company incubators, and for other knowledge institutions, include 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 related to the organization's performance. The network components are displayed in Fig. 2.

Each network component is in itself a system made up of a set of knowledge repositories and intelligent software agents. For instance, research products component is a publications repository which follows the taxonomy shown in Fig. 3. Publications may be refereed or not; if refereed, they may or may not be specialized, and, if they are specialized publications, they may include articles appearing in indexed journals or in conference papers. Some publications are inventions, such as patents, utility models, industrial designs, industrial secrets, trademarks, or technical reports. Divulge publications such as white papers, may have a less strict review process, which is also true of the various types of books shown in Fig. 3.

The value associated with 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 is based on the prestige of the journal, determined by its *impact factor*. The impact factor depends on the number of citations the articles receives from articles in other journals (Garfield, 1999). University rankings as well as faculty awards rely strongly on this measurement. On the other hand, a publication's economic value, for instance, a patent, stems from 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 21st century face the challenge of making these two views of publications coexist in a sustainable manner (Etzkowitz, 2001).

Human resources constitute the intellectual capital of any knowledge organization. In our approach, the key elements are the people who generate knowledge, namely researchers. Researchers may include research professors, full time researchers, such as postdoctoral researchers or research assistants holding Ph.D. degrees, and research students of various levels. Research staff consists of those holding managerial positions, including

5 http://en.wikipedia.org/wiki/Object-relational_mapping.
6 http://www4.wiwiss.fu-berlin.de/bizer/D2RQ/spec/.

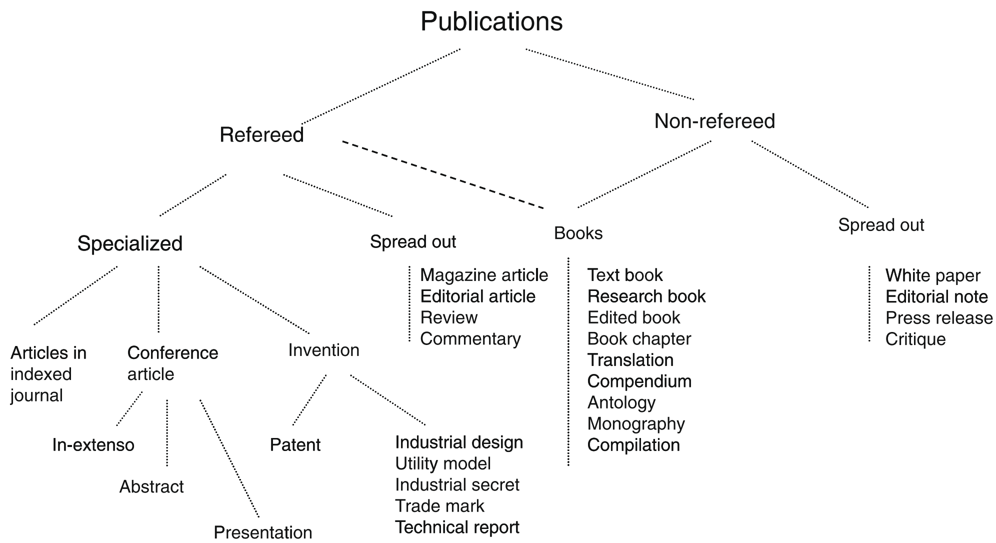

Fig. 3. Research products: publication taxonomy.

provosts, deans or chairmen, technicians at laboratories or experimental sites, and administrative staff. Fig. 4 shows the 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. Fig. 5 displays the 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 with access regulated by a set of software agents. For instance, the publications repository stores publication attributes and the document files. A journal article entry would have the title, authors, journal, dates, citations, impact factors and other features of the article. For a patent, title, inventors, status, countries, licensing contracts, clients, technology claims and other attributes are given. 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. In the case of human resources repositories, research professors are described by a set of attributes with all the information necessary to generate a professor's curriculum vitae. The same is true of researchers and research staff. For research students, their progress and performance is tracked and the student repository stores their transcripts, theses, projects and publications for both current students and alumni. This is a kind of registrar at the 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 are formally defined and stored. An academic curriculum includes name, director, course outline, past and current students, faculty teaching the course, and program performance statistics in terms of graduation rates, accreditations, rankings and other features. The same is true of research units, such as a department, a center or a business incubator. For laboratories and experimental units, the equipment available, its capacity, maintenance and policies for use are part of the repository.

3.3. Knowledge and information network users

There are various kinds of knowledge and information users. Among the internal users are faculty, students and research staff, as well as those who hold executive or managerial roles within

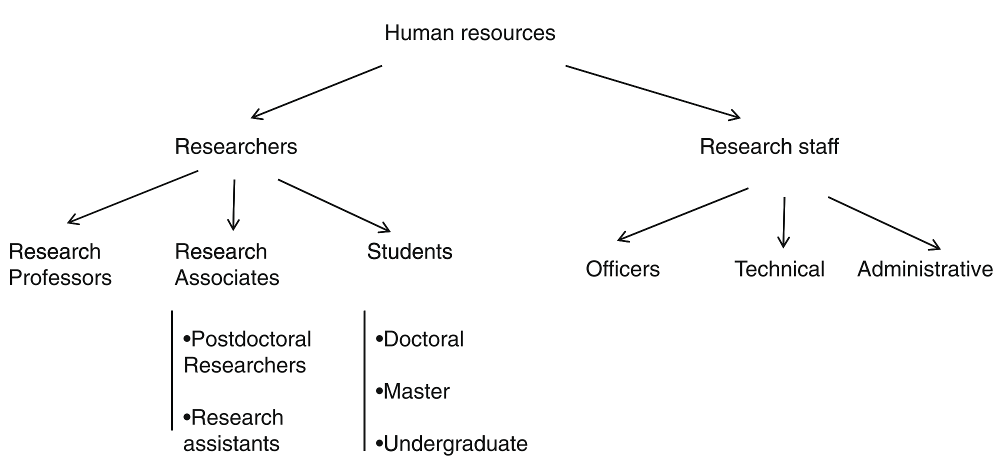

Fig. 4. Intellectual capital: human resources taxonomy.

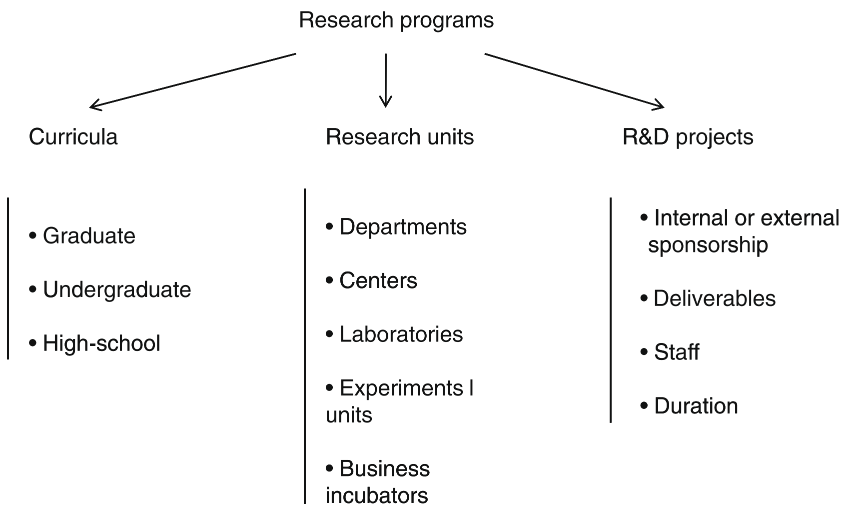

Fig. 5. Research program taxonomy.

the various hierarchical levels of the organization. The KIN system learns these roles and defines views of the corporate memory, based on the user's 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 reports and statistics on the school's faculty, students, publications, etc., but access to information about 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. External users may be prospective students in search of a school, an investor looking for technologies to invest in, a company that is in searching for technologies to license from the university, or a ranking or accreditation agency looking for data to calculate its scores. The user may also be a search engine launched through the internet to gather data for automated business intelligence or benchmarking procedures. Entities communicate among themselves through the intelligent agents that regulate data flows to answer queries or generate reports. This is depicted in Fig. 6. More details of the knowledge and information system are given in (Cantú, Ceballos, Mora, & Escoffié, 2005).

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

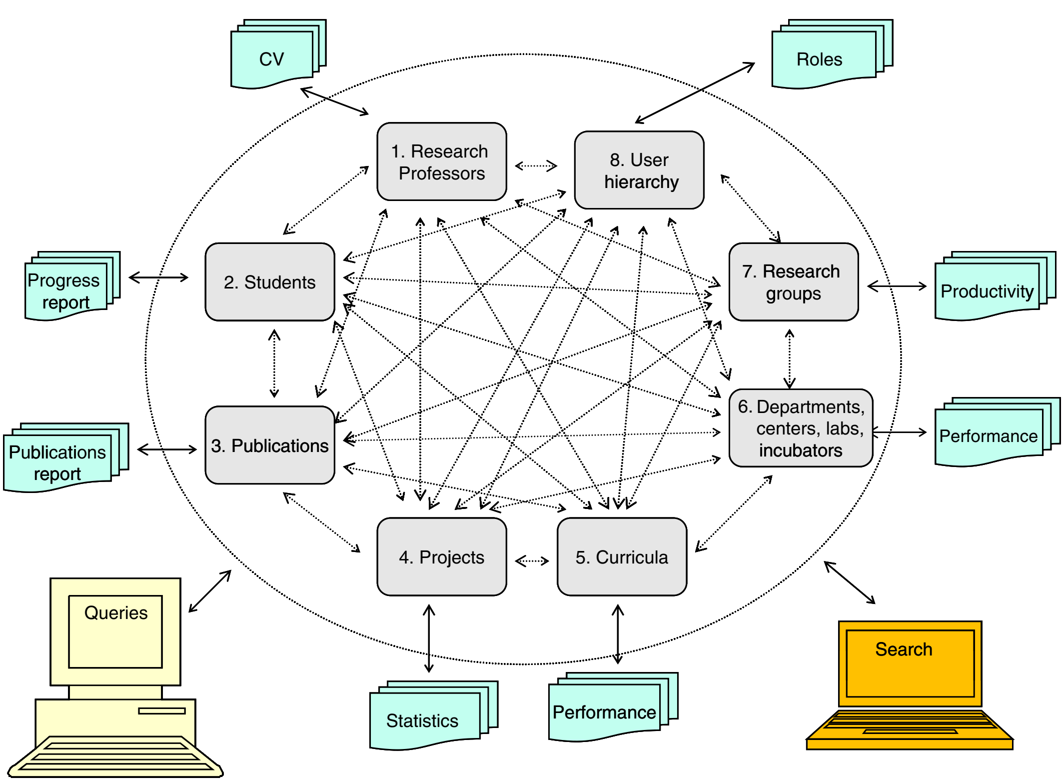

Fig. 6. Knowledge and information network repositories.

Section 3.3. Among these are the "user" who solicits reports or makes 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 Fig. 7.

The KIN platform is made up of four main components: an information system, a multiagent system, a knowledge management system (KMS) and the knowledge-to-information and information-to-knowledge interpreter interface (*K–I* interpreter). This architecture is shown in Fig. 8.

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 in the repositories displayed in Fig. 6. A Model-View-Controller (MVC) framework is used to provide the web interfaces for managing information based on a hierarchical permission system. Web services allow for displaying information from the different repositories from a user-customized perspective; other institutional web portals or applications use the information provided by those web services. The current implementation of the information system is built with the PHP programming language, the relational database management system (RDBMS) is obtained through MySQL, and Mojavi libraries are used for implementing the Model View Controller framework (MVC) (Cantú et al., 2005).

4.2. Multiagent system architecture

The multiagent system is designed to perform off-line operations in the background of the KIN platform (Ceballos & Cantú, 2008).The multiagent system is organized in 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 in the information system databases. 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 Fig. 9.

To perform their tasks, agents execute 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, as needed, during the planning process. Agents use their plans to do several things: to answer user requests, to provide web services, to perform data mining to extract 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 that are responsible for providing services through the instantiation and coordination of other agents, and *utility* agents that 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 the satisfaction of which must be maximized or minimized depending on the task. Information stored in the repository, as well as actions occurring in the information system, are used to generate concrete tasks that the agent delegates to utility agents within a plan. Should a utility agent be necessary for performing a task and that agent does not exist, the main agent is capable of creating it by instantiating its type, based on a library of agent classes previously defined. The creator–created relationship establishes a hierarchy of agents that speed up problem resolution through decomposition of the problem and delegation of concrete subtasks. A utility agent obeys 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 & Cantú, 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 to convert knowledge into information and vice versa in responding to user queries. It is 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 to access the information through the view built on such external ontology.

At the repository management layer there is a set of agents for managing each of the database repositories: the guardian agent that maintains the consistency of the information stored in the repository and provides user services by overseeing the work of

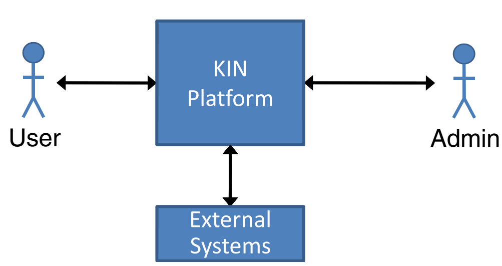

Fig. 7. Knowledge and information network platform.

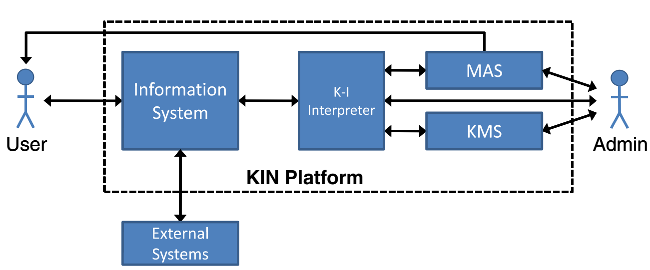

Fig. 8. Knowledge and information network platform architecture.

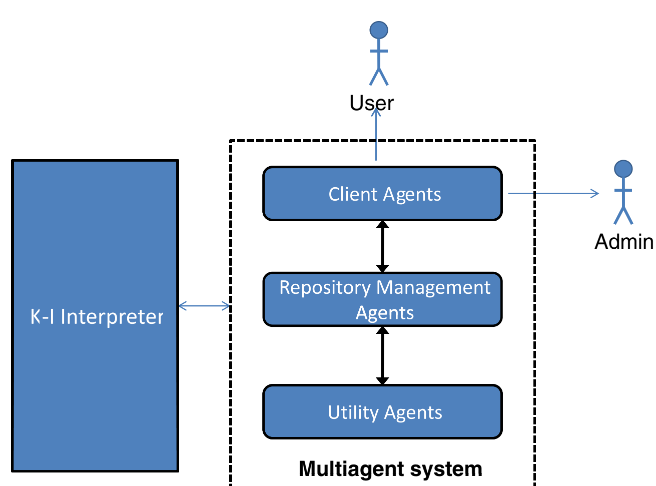

Fig. 9. Multiagent system architecture.

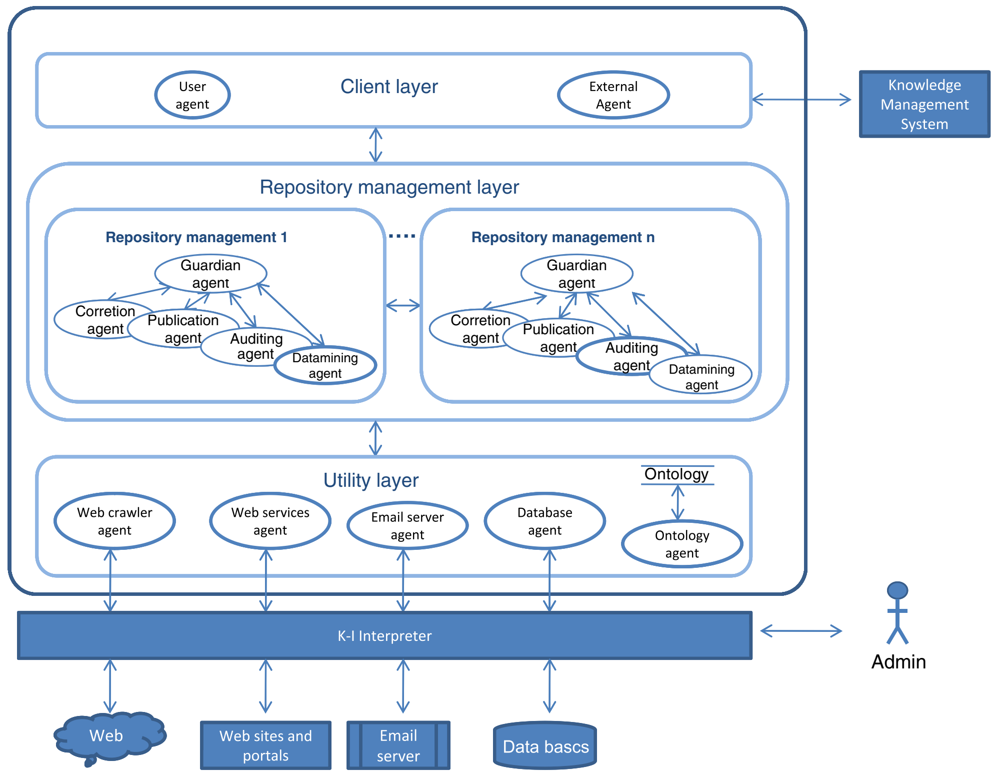

Fig. 10. Types of agents at each MAS layer.

various agents: for instance, the correction agent, which makes automatic corrections on the information; the publication agent, which displays information from certain repositories through web services interfaces; the auditing agent, which conducts system validation; and the subscriber agent, which alerts users to subscription updates. The agents at this layer are capable of doing Bayesian learning during operation to improve their performance in executing their own tasks (Cantú, Garza, Robles, & Morales-Menendez, 2006; Robles, Cantu, & Morales-Menendez, 2005). In addition, there are datamining agents which extract patterns and knowledge from database repositories (Carrillo, Cantu, Garza, & Morales-Menendez, 2005; Gutierrez, Lerma, Salgado, & Cantu, 2002; Rios & Cantú, 2006).

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 & Brena, 2004; Gottlob & Koch 2004). The agents at each of the three layers of the multiagent system are displayed in Fig. 10.

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

The Web Ontology Language, OWL,[7] is used to represent classes of agents, properties and the 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 obtained through various mechanisms, which include a domain ontology manager, information-knowledge mappings management, and a data-source manager, among others. The KMS is composed of a set of ontological repositories and a tool to manage them. 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 Figs. 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

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

8 http://protege.stanford.edu/.
9 http://www.w3.org/RDF.

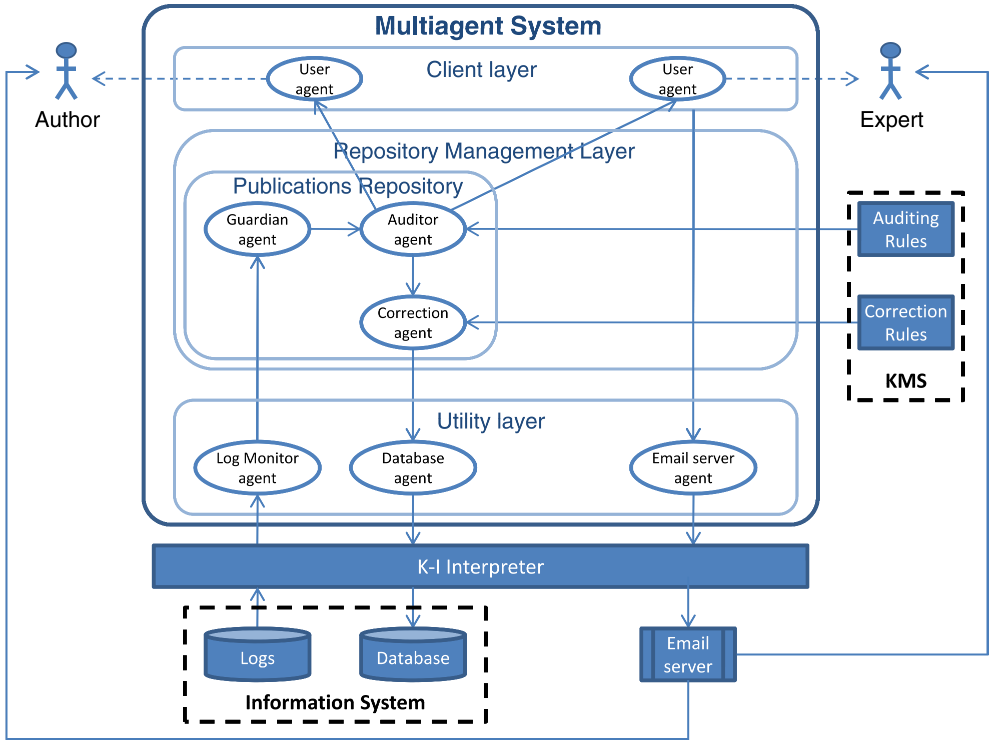

Fig. 11. Multiagent system scenario for auditing publications.

defined by external users for extracting information from the KIN repositories through mappings to the Public RCMO. This architecture is shown in Fig. 12.

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 relate the inconsistency to a correction action, like replacing or setting some attribute of the record or deleting the record itself from the database. Similarly, each evaluation of auditing and correction rules is stored, making it possible to determine probabilistically how effective the rules are in a given case. Cases are codified using attributes contained in 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 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 a mapping to enable utility agents accessing data source repositories through user queries. In contrast, utility agents use parameterized commands from a library for updating, deleting and inserting information in the data sources. In addition, external data and information from external sources are mapped into terms of the RCMO ontology. This is shown in Fig. 13.

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 used for maintaining consistency within the knowledge repositories are of this type. Initially, consistency rules are supplied by the 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 of the auditing result by an agent email and can accept, modify or reject the changes. Inconsistency detection is based on stored rules that describe inconsistency patterns. Fig. 14 shows a causal network of the semi-supervised process for auditing with three possible outcomes. First, if there is a correction available, the correction is made online automatically by applying an available rule. Second, if the inconsistency is not identified or if confidence in 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 of a potential inconsistency. For the first two cases, the user is notified 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, confidence in the publication is

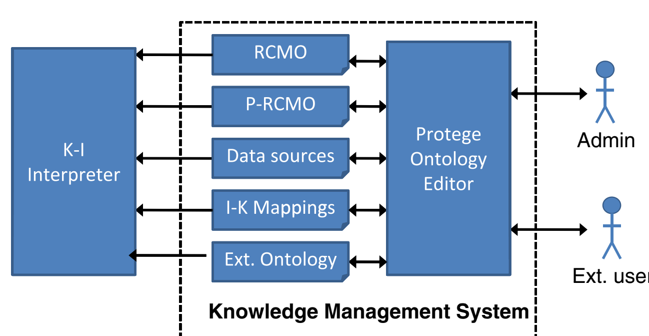

Fig. 12. The knowledge management system.

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

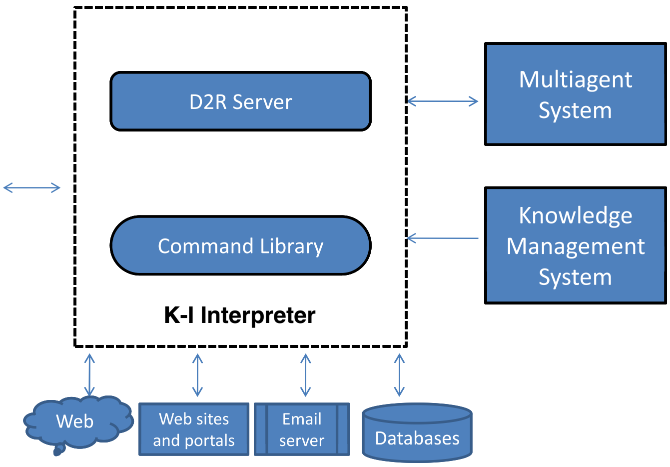

Fig. 13. The knowledge-information interpreter.

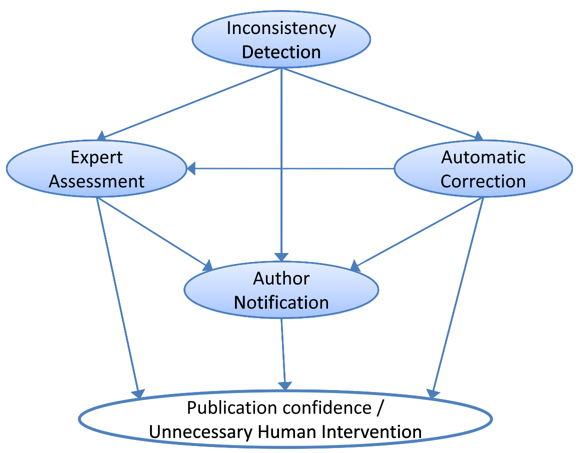

Fig. 14. Causal network of automatic repository consistency maintenance.

calculated in terms of the inconsistencies detected, the corrections made and reinforcement given by experts and users. Audited cases are stored and used for creating parameter learning. In this manner the system modifies its decisions on the basis of its experience (Ceballos & Cantú, 2008).

In this way the multiagent system contributes to the goal of maximizing confidence in the repository by keeping the knowledge repositories consistent by minimizing human intervention. Data mining facilities have also been added to extract knowledge from the database repositories (Rios & Cantú, 2006).

4.6. System implementation

The information system described in Section 4.1 was developed during 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 described in Sections 4.2–4.5 were developed in 2006 and have been operational since 2009. More specifically, auditing and correction rules were modeled using Alloy (Jackson, 2006) and CCalc (Giunchiglia, Lee, Lifschitz, McCain, & Turner, 2004), as described in Ceballos and Brena (2008). The auditing process was modeled using the Electronic Institution formalism (Sierra & Noriega, 1997) (Esteva, de la Cruz, & Sierra, 2002), for which an experimental prototype was implemented (Ceballos, Noriega, & Cantú, 2009).

5. Case study

In this section we describe a case study using Tecnológico de Monterrey (ITESM) as an example of knowledge organization using an action research approach (Reason & Bradbury, 2004). ITESM is a university with 33 campuses in México, 60,000 undergraduate students, 12,000 graduate students, and 8000 faculty members, from which around 1200 are research professors. In 2002, an institutional research program was initiated that would change ITESM's profile from a teaching university to a teaching and research university. To achieve this goal, financial resources were allocated to foster research in the various schools of the university's main campus. Research groups were established and supported with seed funds to allow lower faculty teaching loads, payment of student tuition fees and research assistantships, and coverage of traveling expenses, materials and research infrastructure. The areas of research were defined based on market studies and the expertise of human resources, and included biotechnology, health, mechatronics, nanotechnology, information technologies, telecommunications, and resources such as energy, water, air, forests, etc. A research group, composed of about 18 researchers, includes the principal researcher, adjunct professors, postdoctoral researchers, and doctoral, master and undergraduate researchers. As a result of this initiative, research products proliferated, research students became alumni and the doctoral programs were revitalized (Cantú, Bustani, Molina, & Moreira, 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 that included a computer platform and a corporate memory that would assist high-level administrators 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 research asset management (Aguirre, Brena, & Cantu-Ortiz, 2001).

5.1. The 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 Fig. 1. The network elements were classified, as shown in Fig. 2: knowledge products, human resources and research programs.

For the knowledge products the main tool was the publications management system. The system was built around the concept of a *knowledge pump* corporate memory, which gathers the research products generated by faculty and students in the various research groups and academic programs at the university. The taxonomy displayed in Fig. 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 profit derived from its transfer, as in the case of a patent or any kind of invention, or the initiation of a startup company. Currently, the publications corporate memory stores around 38,000 research products of the types shown in the taxonomy, produced by ITESM researchers since 2000.

For the human resources element the main tool was the system to manage researcher types defined in the taxonomy shown in Fig. 4: professors, associates, assistants, etc. Each type is identified by a set of attributes organized by the system and from which various reports can be generated. For instance, the system is able to generate up-to-dated curriculum vitae for 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 2700 research professors, 3740 research students and 6740 external authors who have produced some kind of publication at ITESM since 2000.

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

The knowledge and information elements interact as shown in Fig. 6. For instance, the information system relates these elements in the following way:

- The system receives information about university human resources by means of an automated online 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 a professor feeds in a journal article, the information attributes of that journal are automatically collected from the journal database in the system.
- This database contains the main journals of the scientific disciplines as well as information about impact factor and other attributes. The journal database is automatically updated when a new article is published in a journal that is not yet included. Every new publication is audited by a human expert to make sure it is correctly classified and that all the relevant information has been fed in.

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

- The system generates automatically predefined reports of various types publications, 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 to update personal and departmental web pages through web services.

Fig. 15. Author association interface.

- 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 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*. As another example, author association with a publication is made through a custom interface. Fig. 15 shows the author association interface displayed over the Journal Article register screen.

The statistical reports generated are more elaborate than mere listing. Two main considerations in their design were (1) generating reports by 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 a user can 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 allowing matrixes to be built that conform to the indicator report. There is also a PHP template that formats the results and creates links for navigation throughout reports. An example of an indicator report is shown in Fig. 16.

5.2. Results

The Knowledge and Information Network System described above has been in operation since August 2004 and has proven to be a useful tool in assisting university high-level administrators and researchers to move ITESM towards becoming a teaching and research university and generating value from the university's knowledge assets. Table 1 shows key parameters of this transition from 2002 to 2009 and the goals for 2015.

Research groups have grown from 20 to 125 since 2002 and are expected to reach around 150 by 2015. Researchers can obtain a national accreditation from the National Council of Science and Technology in 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 to continue to increase in the next five years as well as published books authored by faculty. Patents filed per year were almost non-existent in 2002 whereas the number was 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 Ph.D. students, a key element in any university's research strategy, has multiplied by five and will double in the coming years. The number of technology-based spin off companies incubated within the research groups has had a considerable increase and will continue to expand in the near future. In the TIMES Higher Education World University Ranking ITESM has moved up from number 433 in 2007 to number 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,

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

Fig. 16. Statistical report.

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	50
PhD students	100	520	1000
Technology-based spin offs	1	30	100
TIMES Higher Education-QS ranking	433	338	200
Research funding in USD millions	13	57	100

the latter including research grants and industrial research contracts.

6. Conclusions

We described a knowledge and information network approach to handling research assets in a knowledge organization. We believe that institutions which adopt such an approach obtain multiple benefits from it; for instance, documentation of the research products in a corporate memory, increased awareness of the research assets, and an accurate record 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 for gathering research products and for the distribution of the knowledge generated by the research activities. Researchers are now conscious of which are the important journals and conferences in their disciplines and are teaching and encouraging their students to write their theses and projects reports in the form of scientific papers that are peer-reviewed and criticized during oral presentations at conferences. Professors from technological disciplines are also filing patents derived from their innovations and teaching their students to do the same. Research centers and departments have access to historical reports with statistics on 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 and 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 using the support and infrastructure provided by ITESM's entrepreneur program. At the university level, the system provides useful information about the research areas producing the most publications, the researchers with the highest productivity, the journals and conferences in which researchers publish the most, the proportion of papers written jointly by professors and students, the theses which produce published papers, the publications obtained from research grants, graduate programs which are generating scientific papers, and similar topics. Another benefit of the system is the awareness that external entities will gain about the university's research and innovation capabilities. Among these entities are companies and corporations in need of research services, funding agencies, and prospective students looking for challenging research projects and environments. Among the benefits of such a system, encouragement of the emergence of technology-based spin offs and the contribution to the development of knowledge-based economies is particularly important.

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

7. Related work

The use of a system like ours is no different from what is being done in other parts of the world, as globalization has created similar problems across the globe. However, universities that follow approaches similar to the ones described in this paper are in a better position to contribute to their local economies. Many companies and government agencies have developed corporate memories and knowledge management systems to handle their knowledge assets (Liebowitz, 2004; Liebowitz & Beckman, 1998). 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 scientific publications data bases 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 initiating the research chairs program and for the support and academic freedom to carry out research that lead to the results reported in this article. We also acknowledge the staff of the information technologies offices at both campus and university level for their support during system deployment and operation. We would like to express our appreciation to the deans of schools, department and center heads, and especially to the research chairs principal investigators, adjunct professors, postdoctoral researchers, doctoral researchers, graduate and undergraduate students for using the system and contributing to making it operational. Finally, we acknowledge the valuable suggestions of anonymous referees who contributed to improving and shaping this article.

References

Aguirre, J. L., Brena, R., & Cantu-Ortiz, F. (2001). Multiagent-based knowledge networks. *Expert Systems with Applications an International Journal, 20*, 65–75.
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.
Bizer, C. (2003). D2R MAP – a database to RDF mapping language. In *Proceedings of the twelfth international world wide web conference (WWW2003)*.
Cantú, F., Ceballos, H., Mora, S. P., & Escoffié, M. (2005). A knowledge-based information system for managing research programs and value creation in a university environment. In *Americas conference on information systems – AMCIS, USA (pp. 781–791)*.
Cantú, F., Garza, L., Robles, A., & Morales-Menendez, R. (2006). Learning and using Bayesian networks for diagnosis and user profiling. In *Twelfth Americas conference on information systems – AMCIS, Acapulco, México*.
Cantú, F., Bustani, A, Molina, A., & Moreira, H. (2009). A knowledge-based development model: The research chairs strategy. *Journal of Knowledge Management*, 154–170.
Carrillo, M. A., Cantu, F. J., Garza, L. E., & Morales-Menendez, R. (2005). Learning Bayesian network structures from small datasets using simulated annealing and bayesian score. In *IASTED conference on artificial intelligence applications, Austria*.
Ceballos, H., & 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.

Ceballos, H., & Cantú, F. (2007). Integrating semantic annotations in bayesian causal models. In *20 international workshop on description logics (DL'07) (pp. 527–528)*. Bozen Bolzano University Press.

Ceballos, H., & Cantú, F. (2008). Modeling intelligent agents through causality theory. In *Proceedings of the sixth Mexican international conference on artificial intelligence (Vol. 5, pp. 201–210)*. IEEE Computer Society Press.

Ceballos, H., & Brena, R. (2004). *Finding compromises between local and global ontology querying in multiagent systems*. Germany: Lecture Notes in Computer Science (LNCS). pp. 999–1011.

Ceballos, H., Noriega, P., & Cantú, F. (2009). *Autonomic information auditing through electronic institutions*. IIIA-CSIC. Technical report IIIA-TR-2009-07, Spain, July 2009.

Esteva, M., de la Cruz, D., & Sierra, C. (2002). ISLANDER: An electronic institutions editor. In *First international conference on autonomous agents and multiagent systems, Bologna, July 2002 (pp. 1045–1052)*. ACM Press.

Etzkowitz, H. (2001). The second academic revolution and the rise of the entrepreneurial science. *IEEE Technology and Society, 20*(2).

Garfield, E. (1999). Journal impact factor: A brief review. *Canadian Medical Association Journal, 161*(8), 979–980.

Giunchiglia, E., Lee, J., Lifschitz, V., McCain, N., & Turner, H. (2004). Nonmonotonic causal theories. *Artificial Intelligence, 153*, 49–104.

Gottlob, G., & Koch, C. (2004). Monadic datalog and the expressive power of languages for web information extraction. *Journal of the ACM, 51*(1), 74–113.

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 (Vol. 2313), Springer-Verlag.

Jackson, D. (2006). *Software abstractions: Logic, language and analysis*. MIT Press.

Liebowitz, J. (1999). *Knowledge management handbook*. Boca Raton FL: CRC Press LLC.

Liebowitz, J. (2004). *Addressing the human capital crisis in the federal government: A knowledge management perspective*. Burlington MA: Elsevier.

Liebowitz, J., & Beckman, T. (1998). *Knowledge organizations*. Washington, DC: Saint Lucie Press.

ÓLeary, J., Quacquarelli, N., & Ince, M. (2009). *Top universities guide*. London, UK: Quacquarelly Symonds Ltd.

Padgham, L., & Winikoff, M. (2004). *The Prometheus methodology. Chapter in methodologies and software engineering for agent systems* (Vol. 11). Springer.

Pearl, J. (2000). *Causality: Models, reasoning, and inference*. Cambridge University Press.

Reason, P., & Bradbury, H. (Eds.). (2004). *Handbook of action research: Participative inquiry and practice*. London, UK: SAGE Publications, Ltd.

Rios, M. T., & Cantú, F. J. (2006). Knowledge discovery in academic registrar data bases using source mining: Data and text. In *Twelfth Americas conference on information systems – AMCIS, Acapulco, México*.

Robles, A., Cantu, F. J., & Morales-Menendez, R. (2005). A Bayesian reasoning framework for on-line business information systems. In *Proceedings of the eleventh Americas conference on information systems, Omaha, NE, USA, August 11th–14th*.

Sierra, C., & Noriega, P. (1997). A formal framework for accountable agent interactions. In *Fifth Bar–Ilan symposium on foundations of artificial intelligence, Ramat-Gan, Israel, June 1997 (pp. 23–24)*.

Uschold, M., & Gruninger, M. (1996). Ontologies: Principles, methods and applications. *Knowledge Engineering Review, 11*(2).

Weiss, G. (1999). *Multiagent systems: A modern approach to distributed artificial intelligence*. England: The MIT Press.

Wooldridge, M., Jennings, N., & Kinny, D. (2000). The Gaia methodology for agent-oriented analysis and design. *Journal of Autonomous Agents and Multi-Agent Systems, 3*, 285–312.

Zuckerman, M. (2004). *America's best graduate schools*. US news & world report. USA.