PRO-INOVA
Virtual Platform for Innovation Management in Doctoral Schools

Gheorghe Sebestyen[1], Marius Bulgaru[2] and Laura Grindei[3]
[1]*Department of Computers, Technical University of Cluj-Napoca, G. Baritiu 26, Cluj-Napoca, Romania*
[2]*Department of Manufacture Engineering, Technical University of Cluj-Napoca, G. Baritiu 26, Cluj-Napoca, Romania*
[3]*Department of Electrotechnics and Measurements, Technical University of Cluj-Napoca, B. Munci, Cluj-Napoca, Romania*

Keywords: eLearning, Digital Library, Innovation Management, Document Search, Information Retrieval, Patents.

Abstract: Innovation is an important goal in any doctoral study program. In order to offer support for research and innovative developments we implemented a platform that manages information contained in patent repositories. The platform assures functionalities for information retrieval, content harvesting, patent repository administration and domain specific ontology management. The platform has also an eLearning component that teaches the students the steps of patenting an innovative idea. The platform offers support for gathering and preserving experience for research teams. This experience is benefic mainly for new researchers and PhD students.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

1 INTRODUCTION

Education in PhD schools is mainly based on research and innovation activities. In the initial stage, PhD students must upgrade their knowledge with the latest scientific and technical achievements obtained worldwide on their field of interest.

As the information quantity is growing every year at an exponential rate, it is difficult for the students to retrieve relevant and quality information and knowledge that is needed in their research. In traditional PhD education, the PhD coordinator has a major role in offering the starting documentation necessary for a given research.

However, this documenting process is not formalized and there are few tools that may be used in order to transfer knowledge and experience from a PhD coordinator or research team to a newcomer (Leonard, 1998). Our goal was to develop such a tool that can store the accumulated knowledge of a research team on a given domain of interest.

As shown in the next chapters, knowledge is stored in different forms, from domain ontology (concepts, associations and classifications) toward collections of documents gathered in time by a research team.

In order to promote innovation in technical fields we used as information source international patent repositories (e.g. USPTO, EPO, etc.). The arguments behind this decision were that patented ideas usually have an economic potential and the information inside the patents is highly formalized, enabling computerized information search, extraction and data mining (Heusch, 2006). Digital libraries with scientific articles and eBooks (e.g. IEEE Digital Library) are also important sources for documentation, but usually these repositories have different and restricted access policies (e.g. access or download fee is needed) and they do not offer automated services for information extraction and document downloads.

Our project's main objective is to provide computational means for a new educational program in the field of innovation management. This program is mainly for the PhD schools, but it can be used in any research and development activity where patenting new ideas is an important goal.

The proposed solution is a virtual platform named Pro-Inova with three components:

- **Plat-inova** - a digital library application designed for managing information contained in patent repositories.
- **e-Inova** - an eLearning tool dedicated for teaching innovation and patenting procedures in the frame of an online course entitled "Innovation Management".
- **Demo-inova** – a video tutorial and a user's manual for the Pro-Inova platform.

Sebestyen G., Bulgaru M. and Grindei L..
PRO-INOVA - Virtual Platform for Innovation Management in Doctoral Schools.
DOI: 10.5220/0003898801120117
In *Proceedings of the 4th International Conference on Computer Supported Education* (CSEDU-2012), pages 112-117
ISBN: 978-989-8565-06-8
Copyright © 2012 SCITEPRESS (Science and Technology Publications, Lda.)

We decided to develop a dedicated eLearning tool instead of using a better knowned one (e.g. Moodle, Dotlrn) because in this way we had the opportunity to adapt the on-line teaching sevices to the specific needs of a doctoral school. We also had the possibility to develop interfaces on Romanian language.

2 INFORMATION RETRIEVAL AND KNOWLEDGE REPRESENTATION

In the last years, significant research effort was focused in order to find more efficient methods for information search, retrieval and automated document classification (Grossman 2004). Classic keyword-based document search methods used by most of today's Internet search engines have limited capabilities in expressing the search intention of the users. It is difficult to formulate in a few keywords a given interest in a scientific domain. The semantic content of the search is also affected by synonym and homonym words.

The research in this field is focused in the following directions:

- to find new methods for semantic representation and storage of information (Messerly, 2000)
- to implement efficient algorithms for information retrieval, data mining, and content classification (Fuhr, 1992); (Salton, 1975)
- to develop tools for knowledge management (Fensel, 2002)

In the Plat-Inova component, we tried to integrate some of these research results in a digital library type application. For instance, in order to handle the knowledge related to a given research domain we developed tools for defining thesauri of concepts, relations and associations between concepts. These elements of a domain ontology enable semantic search and classification of patents and information contained in them.

A typical scenario is the case when a research group led by a PhD coordinator expresses its knowledge and experience related to a domain in the form of a particular ontology. Than the concepts and relations in the ontology are used in the search and classification phases in order to enhance these procedures with semantic content. For instance if a user formulates a search expression with a number of keywords, than similar concepts or other more general or more particular terms may be optionally included in the expression. Navigation (browsing) between documents is possible using the association relations between concepts.

We also used the ontological references in order to guide the classification of documents (e.g. patents) based on predefined sets of concepts (thesauri specialized for narrow domains of interest). We implemented procedures that can compute the similarity ratio between a patent and other patents in the repository, or the similarity with a given thesaurus. In this way, the search intention of a user may be expressed through a patent given as an example and not just as a limited number of keywords.

Another form of recording experience and knowledge is through collections. The application has functions for defining collections (individual or collective ones) and associating patents to collections. A given group or individual researcher may define its own collection of patents that reflects a given research interest. A newcomer in the group may start his/her documentation by evaluating the patents in the group's collections or in individual collections of more experienced group members. Collections may be organized on different criteria (e.g. content type, ownership, destination, etc.). The same patent preserved in a single location may be part in a number of collections.

For the same purpose of recording experience, the application stores search queries and expressions. These search strategies can be reused later in time when it is supposed that new patents are released in a direction of interest. PhD students can obtain valuable updated information by reusing search strategies defined by more experienced members of the group.

3 PLAT-INOVA – A TOOL FOR INFORMATION RETRIEVAL

Our solution is based on some of the latest results in the field of digital libraries, information retrieval, content management and semantic classification of information. As information source, we considered patent repositories because of their public availability and because the information contained in patents has a standard format, enabling automated content processing.

In our vision, a research for a PhD thesis, in most cases, is not an individual activity. The PhD student is part of a research group coordinated by a PhD conductor; a number of such groups are organized in a doctoral school and they cover some research domains. The application models this kind of

organization, offering specific services for PhD students, coordinators and simple visitors in order to handle information and knowledge in a number of research directions. Every user has access rights and responsibilities in accordance with their status, experience and interest.

The goal of our platform is twofold: to provide assistance for PhD students involved in research activities and stimulate the innovation process by offering quality information about their domain of interest and to teach basic concept and procedures used in the process of patenting.

The patent harvesting module (presented in Figure 1) is responsible for downloading patents from different sources available on the Internet, based on a given search strategy defined by a user.

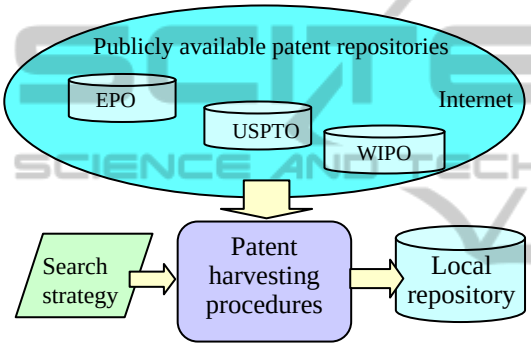

Figure 1: The harvesting module.

This service allows a research group to build its own local repository of patents. The service has a search part, which tries to find patents that fit into the harvesting strategy and a download part that handles the transfer of information and files associated with a patent. As patent sources, the application is using some very known patent repositories such as EPO – European Patent Office, USPTO – United States Patent Office or WIPO World Intellectual Property Organization.

In the case of EPO, the download procedure is made through a web service offered by this repository for automated search and download. In the case of other repositories (e.g. USPTO and WIPO) such a web service is not available and therefore the download is made by simulating an access through the user interface of theses sites. This approach is less reliable because any change in the interface of these sites imposes changes in the download service of the application.

The automated patent downloading is implemented with the OPS v2 Web services that establish a connection with the EPO online patent database. The specifications provided in the harvesting policy are translated into the Common Query Language (CQL) supported by the OPS services and a search query is created. The application sends a SOAP message request containing this query, invoking the bibliographic search service of the OPS.

Since the method presented above only guarantees the download of textual information, we have also integrated an alternative tool for automated patent downloading, using a full-document patent downloader component. This component connects to the European publication server (EPS) via the Web service described in (EPS-WS) and provides the same facilities as the online interface offered by the EPO site.

The request message sent to the EPS server contains the metadata of a patent and a specification regarding the preferred format of the resulting patent. The available choices are PDF files or XML files, with the option of requesting the drawings belonging to the patent, as well.

The PDF format is a scanned version of the patent. Users prefer this format in case of visual inspection of a patent. However, for automated information search and retrieval, this format is useless, because text parsing and search procedures cannot be performed efficiently on scanned documents. For such purposes, the application stores the textual (XML) version of the patent. In such a file tags indicate the different sections of a patent. Using these tags, the search procedure can be directed only to some specific parts of the patent (e.g. abstract, description, claims, etc.).

The EPS Web service (EPS-WS) used for downloading the PDF version offers a limited number of facilities. The EPS-WS only provides patents available in the European publication server (having the “EP” index), similarly to the online interface, which considerably reduces the number of downloadable patents (unlike the OPS services). In order to request a patent through this service, the identification information (namely the patent publication number) must be known prior to defining the request.

Since the EPS-WS does not offer the means to obtain the required information, this component is used in combination with the OPS services that include the bibliographic search service. The list of identification information obtained through the OPS services are used in requests for both types of services (OPS and EPS-WS), obtaining copies of the same “EP”-type patent in text and PDF format.

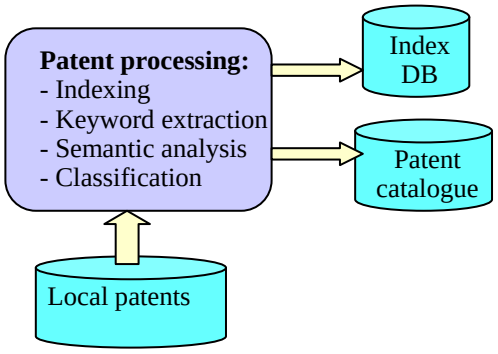

Figure 2: The patent processing module.

The Patent Processing Module. The patents downloaded through the patent harvesting process are used in the next step as the input data for the patent processing module. The patent processing procedures are using only the textual content of a patent.

The patent processing module is composed of several procedures, shown in Figure 2, each responsible for implementing a given functionality. In the first stage of the processing, patents are parsed in order to extract relevant keywords and to eliminate stop-words (words that have only a syntactic role and no relevance for the content). We also disregard words that appear in all documents or in a single document since they provide no semantically useful information for the search process.

The patent indexing procedure was built on the Apache Lucene (Lucene) text search engine library. The goal is to generate an index reference file over the local patent repository in order to increase the speed of the search. Each section of a patent is indexed separately allowing keyword search on different sections of a patent. In this way, we can implement advanced search procedures similar to those present in international patent repositories through combination of conditions on different sections. An extra facility, not present in other patent management applications, is the search in the description and in the claim section of the patent.

The indexing results play an important role in computing the "relevance" score of a patent in a search result list. The score indicates how similar a given patent is with the search expression. The resulting list of patents is ordered based on the relevance score. Again, the Lucene library offers some useful routines for the score computation.

For semantic processing of patents' content, we implemented ***the semantic analysis module*** that works according with the Latent semantic indexing (Deerwester, 1988) algorithm.

The search module contains a number of services (Figure 3) that rely on the data structures created and managed by the processing module.

The searching module is designed to support index or keyword based searches, semantic searches and it is planned to be extended with a full-document searching facility. The semantic search is using the information contained in the thesauri of a domain. The user can determine the similarity of a patent with a given thesaurus or with other patents.

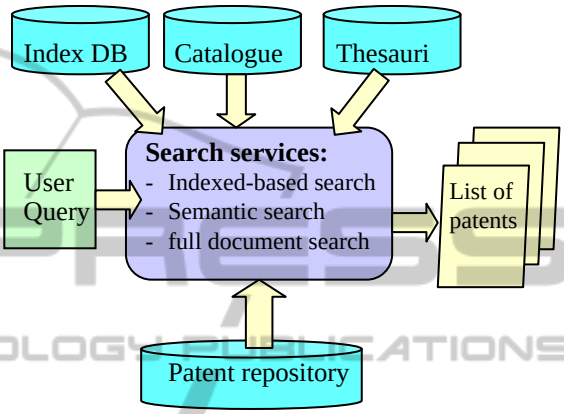

Figure 3: The searching module.

A query may be formulated in terms of a keywords expression that applies to different sections of patents. It may also contain restrictions that apply to the metadata associated with the patent (e.g. period of publication, source of the patent, IPC code, etc.). As a response to a user query, this module returns a list of relevant patents ordered with their relevance score.

The report generation module is responsible for extracting relevant information from a set of predefined patents in order to generate a synthetic search report. The user can specify a report template that indicates which sections and what type of patent information should be compiled into the resulting report.

The report generation module is using the XML format of a patent in order to access directly the different sections of a patent, which will be included in the report.

The ***collection administration module*** implements services needed for the creation, editing and storing of patent collections. The user can select patents from a search result list, which patents to be included into a personal or collective collection. From an implementation point of view, collections are lists of links to patent objects.

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

4 e-INOVA – eLEARNING PLATFORM

e-Inova is the second important component of the ProInova platform, and its main objective is to ensure a specific eLearning environment for the PhD students who wish to learn on-line in the field of Innovation Management.

This eLearning platform consists of several components:

- **eLearning Content Module** designed to offer on line modules of the Innovation Management course including a glossary of terms and a wizard for obtaining a patent in four national and international patent offices;
- **Collaboration Module** with forum and messenger;
- **Content Management Module** for on-line editing and administration of the entire eLearning course, administration of the glossary, creation of the evaluation tests, administration of students marks, etc.;
- **Evaluation Management Module** consists in components for PhD students' auto-evaluation and final evaluation for obtaining a certificate in the field of Innovation Management that is offered on request.

Figure 4: Innovation Management course home page.

e-Inova was implemented in PHP language using web services for integration with Pro-Inova and Plat-Inova.

The eLearning Content Module (figure 4) offers several options for content navigation, links to chapters, modules objectives, glossary, wizard for patents, FAQ, etc.

The Innovation Management eLearning course is divided in modules and each module consists in several chapters and sub-chapters. The text of the chapters includes links to the glossary of terms, and each chapter is concluded with references, links to useful web sites.

The whole course includes also links to the other components of Pro-Inova platform , such as Plat-Inova and Demo-Inova. The content of the course includes the following nine modules:

- Creativity in the technical domain
- Intelectual property
- How to innovate?
- Software for innovation
- Applications, patents in processing of gearing
- Applications, patents in splinting material machining
- Applications, patents in Rapid Prototyping
- Applications, patents in quality engineering
- The history of Romanian creativity

The chapters can be selected using links and the navigation through the entire course site is simple and presented explicitly.

The auto-evaluation tests conclude each module and consist in a set of very simple questions with true/false answers and suggestions regarding the correct answers.

The certification tests are available only for those PhD students who register themselves for this test and wish to obtain a certificate for the course. These tests are more complex, and consist in a theoretical test that can be generated by tutor selecting questions from a database and assigning them to each student.

The new terms in the innovation domain can be added in the glossary of terms and can be accessed directly from the content by links or in the glossary by term, definition or both.

The collaboration module includes two components:

- the forum component that allow users to post public messages and
- the message component for the exchange of private messages.

The Content management module consists in several components designed for adding, editing and erasing each unit of course content, glossary administration (add, edit, erase terms), management of students' evaluation , etc.

All these facilities are available for the tutor and administrator of the course. The content management module was implemented as a database using MySQL and PHP.

The content adding/editing interface integrates a WYSIWYG application that allow tutor to format and align the text (type of fonts, italics, underlined, bold, bullets, numbering, etc), to include images, hyperlinks, to insert date and time, etc directly in the content .

Tutors can change background and fonts colours as well. Modules, chapters and subchapters can be renamed, modified or deleted.

The evaluation management module allows tutor to configure auto-evaluation and certification tests . The theoretical component of the certification tests can be generated by the tutor selecting the questions that are already added in the database giving the opportunity to create new tests for new PhD students very easily.The practical test consists in filling the forms for obtaining a patent from one of the four national or international organizations (OSIM, EPO, WIPO, USPTO)

5 CONCLUSIONS

In order to teach and guide PhD students in their research and innovation activity we developed a computer-based platform that offers the means for information retrieval, knowledge and experience storage and training on patenting procedures.

The platform includes a Plat-Inova application that handles information contained in patents. The application offers services for patent harvesting from different international repositories, for processing of patent content and for efficient retrieval of information. An important goal was to develop some functions that facilitate the preservation of experience and knowledge inside of a research group. Thesauri definitions, selective patent downloading and storage, collections and query expressions administration are some examples of tools used for this purpose.

The E-Inova component of the Pro-Inova platform is an eLearning tool destined for training PhD student in the field of patenting and innovation. The application provides useful information regarding the process of innovation and the formal steps necessary to protect the intellectual rights. It contains a set of interactive tests that allow certification of students in this area.

ACKNOWLEDGEMENTS

This work was supported by the European Social Fund through Sectorial Operational Program for the Developemnt of Human Resources, grant code POSDRU/21/1.5/G/24239.

REFERENCES

Fensel, D., (2002) Ontology-based knowledge management, *Computers, vol. 35, Issue 11*, IEEE Computer Society

Grossman, D. A. and Frieder, O., (2004) Information Retrieval: Algorithms and Heuristics, *The Information Retrieval Series*, Retrieved from: http://www.amazon.com/David-A.-Grossman/e/B001KHBSIU

Salton, G., Wong, A. and Yang, C. S., (1975). A vector space model for automatic indexing, *Communications of the ACM Vol. 18 issue 11*

Fuhr, N., (1992) Probabilistic Models in Information Retrieval. *The Computer Journal, number 3 Issue 35*

Deerwester, S., Dumais, S., and al., (1988) Computer information retrieval using latent semantic structure“, *US Patent 4,839,853*.

Heusch, Ch., (2006) Contributions regarding the processing of the structure and content of patents and development of special system and program, with application for gears, PhD thesis.

Leonard, D. and Sensiper, S., (1998) Divergence and convergence in thinking - "The Role of Tacit Knowledge in Group Innovation", *California Management Review, Volume 40 issue 3*

Messerly, J. Heidorn, G. and Dolan, W., (2000) Information retrieval using semantic representation of text, *US patent 6076051*, USPTO