Exploitation of Informational Resources within Knowledge Society: Digital Library

By Angela Repanovici

I. Knowledge Society

Introduction

A new society, the knowledge society is arising, the elements of this emerging society coexisting with the constitutive elements of the industrial and postindustrial society. The most obvious aspect of the new society is the speed of use, application and dissemination of the communication and information technologies, which puts in the shade the fact that there occurs a major transformation of concepts, structures and institutions specific to the previous society. This insufficient correlation between the evolution of the information and of other components of the emerging knowledge society has led to an approach to knowledge society issues dealing with either the novelty of information and communication systems or the spiritual aspects related to it; it also maintains the confusion among the various names given to the new society, named either "post-modernist", "post-industrial", "informational" or "knowledge" society. The sociologist A. Giddens, considering that the new communication and information technologies have contributed to the wide world spreading of the western culture, under the form of democracy and market economy values, defines the present period as "advanced modernity" not as post-modernism. He states that modern institutions differ from all previous forms of social order only "due to their dynamism, to the extent to which they undermine the traditional habits and customs and due to their global impact".

In fact, the reorganizations we are witness to suggest the fact that the transformations of the current period are equivalent with a radical paradigm change in the sense of the one defined by T. Kuhn in his classical work. Ever since 1993 Peter Drucker stated: "We can be sure that the world arising from the current order of values, beliefs, economic and social structures, concepts and political systems, in other words the order of conceptions of the world, will be different from what one would imagine. In some areas – and especially within the society and its structure – basic transformations have been already developed". [37]

The term of knowledge-society is nowadays used across the whole world. This name is an abbreviation from the structure knowledge-based society.

The cognition is the process through which there is obtained a total new side of knowledge, no matter by whom, human, animal, live machine, people network, humans and machines, and by all forms of organization that can bear knowledge. The cognitive science will get extended not only in order to integrate all these types of cognition, but also integrative, as the cognition will imply integrative processes (structural-phenomenological).

Knowledge, economic factor

A new aspect of knowledge is that of economic factor. In the last 500 years, Laurence Prusak notices, the factors of production were land, labor and capital goods, being neglected the role of knowledge as distinct factor of production. To Prusak, the knowledge represents an intellectual capital, the things learnt by an organization ["there is no sustainable advantage but what a company knows, how it can use what it knows and how fast it can learn something new".

"I don't believe the current theory of knowledge may face these new approaches". [31, 32]

Richard W. Everett (Chase Manhattan Bank, New York City) notices:

"Many economists have argued that technological progress is really nothing but quality improvement in human beings. Some economists take even a broader view and speak of the 'production of knowledge' as the clue to technological progress. The production of knowledge is a broad category including outlays on all forms of education, on basic research, and on the more applied type of research associated especially with industry … invention and innovation."

Roger E. Bohn remarks:

"Philosophers have analyzed the nature of knowledge for millennia; in the past half-century, cognitive and computer scientists have pursued it with increased vigor. But it has turned out that information is much easier to store, describe, and manipulate than is *knowledge*."

This author shows that it is important to understand the **technological knowledge**, i.e. the knowledge about the best way to produce goods and services. Bohn, like others, makes a distinction between data and information.

Similarly, he makes a distinction between information and knowledge. **Data** emerge directly from measuring one or more variables. The items of information stand for data that have been organized or structured in a certain way, placed in a context and having a meaning. The **information** shows, generally or partly, the condition of the production system. **Knowledge** is much more. It pursues to understand the process, to produce causal associations, to make predictions, to make prescriptive decisions.

It is to retain Bohn's definition for the notion of learning: "Learning is evolution of knowledge over time".

Both define **technical knowledge** as understanding of the effects the input variables (x) have upon the output variables (Y). As Y = f(x), the technological knowledge is the knowledge upon arguments and behavior of f(x). The author identifies eight stages of technological knowledge. The more advanced the technological knowledge is the closer to science it is and it can, therefore, be formally managed.

The stages of technological knowledge are as follows:

- Complete ignorance of the nature of process;
- Acknowledging the process. Analogies with other processes. Enhancing knowledge outside the organization. This knowledge is tacit and placed in workers' mind. Production is more than an art. Problems are worked out through trial and error.
- Measuring stage. Variables may be measured but are yet to be controlled. This stage is preparing the following stage. It is the pre-technological stage, the knowledge is propositional, written and also unwritten.
- Variables control, but not with great precision, at a medium level. The knowledge is written or/and comprised in hardware. The learning process takes place through experiments, with scientific method. The organization is mechanically-based.
- Variables may be controlled with precision for a large range of values. There are used operation guidebooks. Workers' role changes: the focus is now on problem solving.
- Characterization and identification of processes. It is known the way variables affect the results if reduced variations of variables are produced. The process can be slightly adjusted, there can be produced reaction systems (feedback) for control. The problem solving process can be carried out through experimental scientific methods guided by appropriate theories and simulations. Learning and improving are now developing. The technological knowledge is found in

databases and software. The organization focuses on learning.
- The stage of scientific level that answers the question why. It is now outlined the scientific model of the process and the way it operates at a wide scale so that it includes non-linear and interaction effects of some variables with others. The process may be optimized and the process control may be automated. This stage is called the automation stage.

G Anthony Siesfield remarks that **knowledge cannot be measured but only its effects.** Reconsidering L. Pusak's foregoing idea, he shows that the idea of knowledge was not outlined as a stock but as a flow and in this very flow there are interwoven people's experience and inspiration leading to knowledge applied in technological processes and business management.

Ikujiro Nonaka makes some very interesting observations:

> " few managers grasp the true nature of the **knowledge-creating company,** let alone know how to manage it … The reason: they misunderstand what knowledge is and what companies must to do to exploit it. […] A company is not a machine but a living organism. Much like an individual, it can have a collective sense of identity and fundamental purpose. This is the organizational equivalent of self-knowledge- a shared understanding of what the company stands for, where it is going, what kind of world it wants to live in, and most important, how to make that world a reality. In this respect, the knowledge creating company is as much about ideals as it is about ideas. And that fact fuels innovation."

In such a company the conception of advanced knowledge is not a specialized activity of the department of research and development. It is a manner of behavior, a manner of being. In such a company each one is a knowledge worker, which confers it the entrepreneur character.

Giovanni Dosi considers the economy as a distributed system of knowledge objects. This author makes the following distinction between information and knowledge:

> " The **former** entails well-stated and codified propositions about state of the world (e.g., it is raining), properties of nature (e.g., A causes B) or explicit algorithms on how to do things. On the other hand, **knowledge**, in the definition I am proposing here, includes: i) cognitive categories; ii) codes of the interpretation of information itself; iii) tacit skills; iv) problem-solving and search heuristics irreductible to well-defined algorithms." [30]

Dosi underlines the fact that "in modern economies, firms are major, albeit by no means unique, repositories of knowledge. Individual organizations embody specific ways of solving problems that are often very difficult to duplicate in other organizations or even within the organization itself. In turn, **organizational knowledge** is stored to a large extent in the operating procedures (the routines) and the higher level rules (concerning what to do when something goes wrong or how to change lower level routines) that firms enact while handling their problem-solving tasks in the domains of production, research, marketing etc."

Dalke Neef synthesizes the role of organizational and technological knowledge: "In the knowledge-based economy it is the production of ideas, not goods, that is the source of economic growth, and the reason that the new computing and telecommunications technologies are so economically revolutionary in their nature is that they allow ideas – in the forms of technics, research results, diagrams, drawings, protocols, project plans, chemical formulae, marketing patterns, etc, - to be distributed instantaneously and in a coherent way to anyone, anywhere in the world".

The complexity of information not only represents the quantity of information but also makes reference to interactivity, i.e. the extent to which it can be especially elaborated for the receiver, in line with the correctness or security of information. The complexity of information means in fact the quality of information.

Acquiring information means the number of people that can receive a certain item of information.

The different conceptualizations of the term of informational society lay on different dichotomist sets: postindustrial society vs. informational society; mass communication vs. new communication technologies; capital goods and labor vs. knowledge, information, and communication. Beyond the variety of analytical frames, most definitions associate the informational society with a series of society-based transformations that would mark the transition to a new type of society that has become dependant on complex electronic information and on communication networks; a society allotting a significant part of its resources to information and communication-oriented activities.

Wurster and Evans notice that as long as the information is comprised within delivered physical objects, there is a basic law governing its economy: the higher the information complexity the lower the possibility to acquire it.

The new conception on knowledge capable of rendering both material and transcendental reality has been associated by some theoreticians with the technology, considered not only a discipline but also active and effective information [27]. Technology is generally described as industrially useful knowledge or as the process of acquiring this type of knowledge concretized under the form of invention, innovation, dissemination, knowledge protection, economic production, know-how, including continuities and discontinuities in information production and dissemination or the imitation and new technical solutions. However, it is considered that technology, as well as knowledge, have a transcendental dimension, a "realm of ideas", being connected to systems of beliefs, values and preferences associated with them, which inhibits or validates certain actions not others [26].

The changes of approach to the organization's objectives in the information era (a new form to present the new economy) as compared to the industrial era are also numerous: the mass production becomes mass customizing, the mass marketing is replaced by the individual marketing, the optimization of the information chain and the informational collaboration with providers replace the optimization of the physical chain and of the physical collaboration with providers, the virtual globalization is more important than the physical location, etc.

The new economy has generated a new type of labor, the information-based labor and a new type of worker: the information worker. He is characterized by the capacity to reason and to know at a high level of education, experience, personality and motivation, different from the worker of the industrial era.

There is prefigured the turning of the information worker into the knowledge worker. Whereas the information worker produces, processes, stores, transmits and compares information, the knowledge worker is to design information producing value, i.e. knowledge.

Knowledge society

Knowledge society represents more than informational society and than computing society, interweaving the both of them.

From the moment the Internet interferes with its great advantages (e-mail, electronic commerce and electronic transactions, Internet market, "content distribution") by comprising a great number of individuals within the electronic information area there occurs the transition to the information society.

The knowledge is meaningful meaning and active information. That is why the knowledge society cannot develop but grafted on the informational society. Likewise, it is more than informational society on account of the major role attributed to the information-knowledge in the society. The most appropriate meaning of the knowledge society is that of information and knowledge society.

The term of knowledge-society is nowadays used across the whole world. This name is an abbreviation from the structure knowledge-based society. Romano Prodi, the president of the European Commission, sometimes uses the syntagm "knowledge-based economy".

If one is searching the theme "knowledge society" on the Internet he/she will find thousands of references. In 2001 the magazine DEUTSCHLAND dedicated a special issue to the knowledge society. Nico Stehr remarks:

"The social order shaped in the horizon is based on knowledge"

The volume of knowledge that is at our disposal doubles every five years. If we wonder about the effect of the current transition from the industrial society to the knowledge society upon the workforce and upon companies, upon politics and democracy – briefly, upon our organizational principles regarding the way we lead our life, then we are entitled to talk about the way we are going to live in the knowledge society.

The knowledge era is working […] Knowledge is the main characteristic of the future societies. […]

The knowledge can be defined as the capacity to act, as an action potential. The scientific and technical knowledge is nothing but the ability to take action. The privileged status of the scientific and technical knowledge in the modern society derives not from the fact that the scientific discoveries are generally considered objective, trustful, and realistic but from the fact that this form of knowledge, more than any other, continually creates new action opportunities. […]

The scientific interpretation must reach "a conclusion" – in order to have a factual value. In the current modern society, this task of turning thoughts into conclusion and of making the scientific visions useful is due to the knowledge workers. [28]

The knowledge is becoming the basis and the principle guiding the human beings' activity. In other words, we are now organizing the reality according to the knowledge we own. [29]

If the main characteristic of the modern society is the knowledge, then the production, dissemination and design of knowledge cannot avoid the political influence. One of the most important problems we are going to be confronted with in the next decade will be the way of monitoring and controlling knowledge. This will generate the development of new branches of the science policy: knowledge policy. Knowledge policy will adjust the volume of the new knowledge that is rapidly increasing within the society and will influence its development.

This last paragraph from the above citation, which makes reference to knowledge society, recalls the notion of political technology previously introduced by the author of this study:

"The political technology is framed as a field studying the social consequences of the new technologies and is looking through the possible technologies that are likely to be achieved, in order to help the current society evolve towards a superior stage […]. At the same time, the political technology is expected to conduct research into the consequences of the new microelectronic, computing and cybernetic technologies upon the human psychology and, inclusively, upon the society, into the changes that are produced within the structure of the labor force, into the way individuals use the time in production and in their extra productive life, as well as into an entire series of other aspects. The political technology may recommend the adaptation of the society to the new processes. […] The political technology may formulate requirements related to technology and science in order to address the society needs, establishing a series of social functions the technical systems are to fulfill, submitting to research the way these functions may be achieved. That is why it also addresses scientists and designers of technology and technical systems. Therefore, the political technology comprises two significant aspects, one addressing the management way of the society, and the other one the innovation way. It brings together the social and technological undertakings - the way the social undertaking may favor those technologies that contribute to the economic and social progress, but also the way science and technology must gather their efforts to address the great requirements of the developing society."

What is knowledge society?
The knowledge society undertakes:

- (I) An extension and elaboration of the scientific knowledge and of the truth on existence.
- (II) Use and management of knowledge framed as technological and organizational knowledge.
- (III) Generation of new technological knowledge through innovation.
- (IV) Dissemination of knowledge towards all individuals through new means, particularly the Internet and the electronic book and through new learning methods, particularly electronic means (e-learning).

A term that has been often used recently is that of new economy. It is known that in the informational society there is framed the internet economy. In the knowledge society there is framed a new economy including the internet economy. This is the reason why the **new economy is the economy of the knowledge and informational society.**

- (V) Knowledge society represents a new economy where the innovation process (capacity to assimilate and convert the new knowledge in order to create new products and services) becomes crucial.

The innovation in the knowledge society pursues to improve the productivity, not only the classical productivities related to labor and capital goods but also the new productivities related to either natural energetic and material sources or the environment protection. That is why the new economy supposes the encouragement of creation and development of innovating plans with an own knowledge structure.

Such plans can be born through cooperation between companies, universities and governmental or public research institutes (including academic).

In a report of the European Communities Commission from 2001 it is shown that in order to get benefits from the new economy there is required a high-performance Internet interface and adequate structural reforms in society, administration and economy.

The influence of the Internet as market in the informational society and the admittance of the assets/goods importance, especially the knowledge, stands for new characteristics of the new economy. Richard Boulton [12]

characterizes in this way the difference between the old and new economy: in the first one it is the tangible goods that matters, in the second one it is the intangible assets creating value that are important.

The intangible is non-material, difficult to describe and especially difficult to measure.

- (VI) Knowledge society is necessary in order to assure a sustainable society from ecological point of view, since without scientific and technological knowledge it is unlike to produce goods, organizations and economic and technological (even biological) changes necessary to save the humanity from disaster in the XXI-th century,
  Consequently, the knowledge society identifies with the informational and sustainable society. It will be very difficult to find another way for sustainability, beside the knowledge society.

- (VII) Knowledge society has global character and stands for a globalization factor. Through both components, informational and sustainability, the knowledge society will have a global character. The knowledge itself, as well as the information, will have a global character.

- (VIII) Knowledge society will also stand for a new stage in culture, the knowledge culture passing in the forefront, implying every type of knowledge, the artistic, literary etc knowledge inclusively. This way, there will be paved the way for what we have called Society of conscience, of truth, morality and spirit.

In table 1.1 the main features of knowledge society will be synthesized:

Knowledge society presupposes:
I. An expansion and a thorough study of the scientific knowledge and of the truth about existence
II. Use and management of the existent knowledge in guise of technological and organizational knowledge
III. Production of technological knowledge by means of innovation
IV. An unprecedented dissemination of knowledge towards all citizens by new means, resorting with priority to the Internet and to the electronic book, and resorting to methods of learning by electronic procedures (e-learning)

V. Knowledge society stands for a new economy in which the innovation process will play a decisive role. The influence of the Internet as market in informational society and the acknowledgment of the importance of the intangible assets value, especially the knowledge, constitute characteristic features of this new economy.
VI. Knowledge society is fundamentally necessary in order to ensure an ecologically sustainable society.
VII. Knowledge society has a global character and is a globalization factor.
VIII. Knowledge society stands for a new stage of culture.

Table 1.1 Components of knowledge society

If informational society is applied the vectors of knowledge society, even during the development of informational society, it is then possible to gain time. Therefore, in our country, the actions undertaken for knowledge society should unfold simultaneously with those concerned with the passage from the informational underdevelopment stage to the informational development. Not in the first place informational society and afterwards knowledge society, but the pursuit of both objectives should be combined from the very beginning.

This is the only way for us to reach a developed informational and knowledge society.

Vectors of Knowledge Society

Two comprehensive classes of knowledge society vectors have been defined:

- Technological vectors
- Functional vectors

A vector of knowledge society stands for an instrument that transmutes informational society into a knowledge society. In order to make the fist move in knowledge society, it is necessary to launch a minimum number of such vectors. The first such vector consists in the creation of a *"developed" Internet*, which is a technological vector, then *the technology of the electronic book* (technological vector) and the knowledge management (functional vector, endowed with two valences, the first for the economic and organizational functioning of a multinational enterprise, corporation or of society itself, the second for the moral use of knowledge within global society). However, the number of these vectors of knowledge society is much more numerous, every new vector making a step forward for the development of this society.

Technological Vectors of Knowledge Society

We deem the following technological factors to be taken into consideration for knowledge society:

Developed Internet, by geographic expansion, by use of transmission bands up to the largest possible, by passing from the IP4 communication protocol to the IP6 protocol, by including each and every institution, residence and citizen within the network.

Technology of electronic book, which differs from the book on the Internet, although its diffusion is based particularly upon the Internet, but also upon CDs. **Intelligent agents,** which stand for expert systems endowed with artificial intelligence, used for data mining and even for knowledge discovery; the intelligent agents will be made much use of for many functional vectors of knowledge society.

Intelligent environment for human life and activity. **Nanoelectronics**, which will turn into the main physical support for processing the information, as well as for many other functions, belonging both to knowledge society and to conscience society.

Functional Vectors of Knowledge Society

The number of these factors may be great enough, as more and more fields of activity increasingly depend upon knowledge:

- Knowledge management for national and local enterprises, organizations, institutions, administrations.
- Management of moral use of knowledge at a global level.
- Biological, genome knowledge (knowledge of the genome and of the functions the genes structure determines).
- Health care system at a social and individual level.
- Environment protection and ensuring of durable and sustainable society by a specific knowledge management.
- Thorough study of the knowledge upon existence.
- Generation of new technological knowledge.
- Development of a culture of knowledge and innovation.
- An educational system based upon the methods of informational society and of knowledge (e-learning) etc.

General theory of information
Information
Phenomenological
Structural-phenomenological - Integrative - Mental
Structural - Syntactical - semantic (information with its own, internal meaning)
Data
Syntactical structural information with an external meaning, in a certain context
Computing program
Information which acts as syntactical information or as intelligent semantic information

Table 2: General theory of information

Knowledge Management at Global Level

The greatest danger which stands out for knowledge society resides in the utmost extension of knowledge society privatization. An equilibrium must be made up between the economic and the moral use of knowledge. There must be defined the sphere of management for the moral use of scientific knowledge within global society.

The knowledge management at a global level should aim at one of the basic targets of knowledge society, that is the dissemination of information-scientific knowledge on the largest scale possible, free of charge or at an extremely low price. This tendency is checked by another contrary tendency, imposed by the new rules of intellectual property introduced especially during the last 10-15 years.

We may take into consideration the following typology:

- Technological and economic systems (global structures; restructuring and economic networks; state-of-art information and communication technologies; new practices of media communication);
- Institutional changes (societal networks; crisis of modern institutions; decline of communities and of traditional social relations; emergence of new types of community; a new public space);
- Ideological and valor changes (neo-individualism; new cultural values and changes; postmodernist practices in everyday life).

II. Role of the Library – Infodocumentary Institution within Knowledge Society

In the framework of the general direction, informational society of the European Commission, one of the priorities aimed at consists of the amplification of the functioning efficiency of the great cultural effects depositaries by means of state-of-art management and interfacing techniques.

Definitions of the Digital Library

- "New digital libraries are characterized by features which are not possible for traditional libraries, this way the concept of library is expanding considerably beyond the physical limits. They will provide innovative resources and services. An example consists in the ability to interact with the information: rather than placing a reader in front of a table with numbers, digital libraries allow users to perform a selection among various ways of visualizing and operating with the numbers, including the graphic charts they can explore. By means of an extensive use of hypertext links in order to interconnect information, digital libraries allow users to find related digital materials upon a certain subject." [127]
- "Digital libraries stand for organizations which provide the resources, including the specialized staff, for selecting, structuring, offering intellectual access to, interpreting, disseminating, preserving the integrity and assuring the duration in time of the collections of digital works so that these ones should be promptly and economically available for use by a pre-established community or by a set of communities." (Digital Library Federation)
- "Digital libraries represent complex data/information/knowledge (henceforth information) systems which contribute to: the compliance with the information necessities of the users (societies), the supply of information services (scenarios), the organization of information upon useful ways (structures), the administration of the information location (spaces) and the settlement of a communication channel with the users and their agents (channels)" (Edward A. Fox, July 1999, in conformity with 5S Framework)
- The activity of the digital library unfolds in the framework of a complex study location molded by four dimensions: community, technology, services and content" [99]
- "The field of digital libraries deals with the amplification of human civilization by applying digital technology to information problems addressed by institutions such as: libraries, archives, museums, schools, publishing houses and other information agencies. The activity upon digital libraries focuses

upon the integration of the services and upon a better compliance with human necessities, by means of a holistic treatment regardless of interface, location, time, language and system. Although substantial creations have been created for the individuals' exclusive use, we deem shared resources one of the defining characteristic features of libraries. Libraries perform connections between people and information; digital libraries amplify and augment these connections" [100]

- For a conclusive discussion upon the community definitions, approaches and perspectives upon "digital libraries", you may refer to "What do digital Libraries represent? Visions in Competition" [9]
- "The digital library is
  - The collection of services
  - And the collection of information objects
  - Which assist the users in the operation with the information objects
  - And the organization and presentation of the respective objects
  - Available directly or indirectly
  - Via electronic/digital means." [32]
- "The digital library stands for a concept which displays different meanings for different communities. As for the engineering and computing community, the digital library is a metaphor for the recently distributed types of data bases services which manage the non-structured multimedia data. As for the political and business communities, the concept represents a new market place for the world informational resources and services. As for the futurist communities, digital libraries stand for the Well's World Brain manifestation. The perspective taken into consideration is rooted in the tradition of information science." [99]
- "An organized data base, comprising digital information objects in different formats, maintained in order to supply unmediated access to a community of users, which displays the following characteristic features:
  - a comprehensive access means (for example a catalogue) has at its disposal a research and retrieval capacity within the entire data base;
  - there are organized technical procedures by whose means the library management adds and removes objects to/from the data base in conformity with a coherent and accessible policy of the collections." [31]
- "Digital libraries constitute a set of electronic resources and technical abilities associated to the former for the creation, search and use of information. In this respect they stand for an extension and for an amplification of information depositing and of retrieval systems which manipulate digital data in any medium (text, images, sounds; static or dynamic images) and which are present within distributed networks. The content of digital libraries includes data, metadata which describe various aspects of the data (for example representation, creator, possessor, copyrights) and metadata which consist of links or relations with other data or metadata, either internal or external to

the digital library (UCLA-NSF Social Aspects of the Workshops of Digital Library).

- "Digital libraries are built – collected and organized – by a community of users, and their functional abilities support the information necessities and the manners of making use of the information which characterize the respective community. They constitute a component of the communities in which individuals and groups interact with each other, resorting to data, information and resources and knowledge systems. In this respect, they constitute an extension, amplification and integration of a wide range of information institutions, as physical places in which resources are selected, collected, organized, preserved and accessed for the support of a community of users. These information institutions comprise, among others, libraries, museums, archives and schools, however digital libraries extend and serve other community groups, including classrooms, laboratories, hostels and public spaces." (UCLA-NSF Social Aspects of the Workshops of Digital Library)"
- "Systems providing a community of users a coherent access to a vast, organized thesaurus, of information and knowledge. This organization of the information is characterized by the absence of a previous detailed knowledge of the ways of making use of the information. The user's ability to access, reorganize and make use of this thesaurus is enriched by the capacities of digital technology"(adaptation from Interoperability, Evaluation and Research Agenda of Digital Libraries)
- "A library that has been expanded and enhanced by the application of digital technology. Important aspects of the digital library which may be expanded and enhanced include:
  - Collections of the library
  - Organization and management of collections
  - Accession of the library units and processing of the information comprised in these units
  - Communication of the information with respect to these units
  - "The generic name for the federate structures that offer humans both intellectual and physical access to the enormous and in continuous growth networks of information encoded in multimedia digital formats." (Digital Library of the University of Michigan: It's not Your Father's Library – Bill Birmingham)
- "A digital library constitutes a medium of distributed technology which dramatically reduces the boundaries from the creation, dissemination, manipulation, depositing, integration and reuse of the information by individuals and groups." (Edward A. Fox, editor, Source Book upon Digital Libraries)
- "A digital library is a representation readable by a machine of the materials, which can be found in the framework of a university library, together with the organization of the information meant to assist the users who search for

specific information. A service of the digital library stands for an assembly of digital processing, depositing and communication machinery, together with the software necessary to resume, emulate and expand services offered by conventional libraries based upon paper and other material means for collecting, depositing, cataloguing, retrieving and disseminating the information (Edward A. Fox, editor, Source Book upon Digital Libraries)

Glossary/Terms related to Digital Library

(by Peter Graham, Rutgers University Libraries)

- digital archive: a digital library which is intended to be maintained for a long period of time, i.e. longer than an individual's life span and surely longer than the individual technological eras (previously, sometimes, also "digital research library")
- digital preservation: preservation of the artifactual information by the digitization of its image (for example scanning a manuscript page, digitally photographing a vase, or converting the registration of a cylinder under a digital form).
- electronic preservation: preservation of the information which exist under digital form (which means electronic), i.e. techniques associated to refreshing, migrating and assuring the integrity.

Techniques of Digital Preservation

- Refreshment: copying the digital information from a long-lasting depositing medium in another medium of the same type, with not the least change in the bit flow (for example, from a degraded 800 bpi band to a new 800 bpi band, or from an older 5 "floppy to a new 5" floppy)
- "Modified refreshment" means copying in another medium of a partial similarity so that no change should be operated in the bit pattern that is connected to the application and operation system which makes use of the data, for example from a 800 bpi band to a 1600 bpi "square" band with cartouche; from a 5 "to a 3" floppy disk.
- Migration: copying data or converting data, from a technology to another, either hardware or software, maintaining the basic characteristics of the data; as a matter of fact far in the future. (Nowadays, it is known, this final qualifier raises numerous questions.) Example: conversion of XyWrite w/p files in Microsoft Word; conversion of ClarisWorks v3 spreadsheet files in Microsoft Excel v4 files; conversion of binary band images of research multi-punched cards in database format; copying of a 800 bpi band file on a sequential disk file; conversion of a database in DOS Fox Pro in a database Visual Basic for

Windows 95; conversion of a PICT image in a TIFF image; conversion of a ClarisWorks file for Windows v4 w/p in a file Macintosh ClarisWorks v4.

We may provide examples, as those offered above, for cases we know to be required; the issue of the long term preservation consists in adopting measures for subsequent migrations, not knowing what the future technologies will be made of.

- Emulate: from the point of view of the hardware terminology, creation of software for a computer which reproduces in all essential characteristics (as they are traced through the intermediary of the problem which is to be solved) the performance of another computer of a different design. Computers may emulate computers from a previous generation in order to ensure the compatibility, or they may emulate a computer from the generations to come in order to offer a media for the development of the software, while the new computer is still in the manufacture phase.

From the point of view of the preservation terminology, the creation of a software which analyses the software media of a document in such a manner that it should provide an interface between the user and the document which reproduces in a significant manner the characteristic features of the document the way it was created from the original software.

III Exploiting the Digital Library Resources in the Knowledge Society

The digital library provides services based on the new information technologies. The users wish to find information within the shortest time possible.

The collections management, funds designated to acquisitions are oriented towards data bases and on-line services.

In order to evaluate services and to improve the performance indicators within university libraries, managers are confronted with similar problems:

- Types of users accessing the services provided;
- Collections accessed;
- Ways to help digital library services become more efficient when meeting users' requirements.

The library users are interested in: the way in which users access these digital libraries and, due to the environment nature, they can access detailed recordings with regard to the way users navigate through these informational spaces.
In order to exploit and manage great amounts of information in the digital library there is advanced an approach to these data bases through data mining applied in libraries.

Data mining (DM) – general outlines

Data mining (DM), known as knowledge-discovery in data bases (KDD), is the process of automatically searching large volumes of data for patterns (Wikipedia article *"Data mining"*[1]). In order to achieve this, data mining uses statistics and pattern recognition.

Data mining was defined as being "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" but also as being "the science of extracting useful information from large data sets or data bases".

Data mining is a general term and it is used at a large scale of contexts with various meanings.

Used in the technical context of data warehousing and analysis, data mining is a neutral term. However, the term was sometimes used with pejorative meaning, imposing patterns (and, in particular, causal relations) on data, where they did not actually exist. This imposing of non-relevant correlations, which induce errors or attribute-based trivial correlations, is the criticized term from "data dredging" statistics. In a more constraint sense "data dredging" implies data scanning for any relations and, when something is found, there is given an interesting explanation. The problem consists of the fact that, invariably, within the masses of data there arise interesting particular relations. Another danger consists in discovering correlations that do not usually exist. The investments analysts are the most vulnerable ones in this area.

[1] http://en.vwikipedial-org/wiki/Data mining

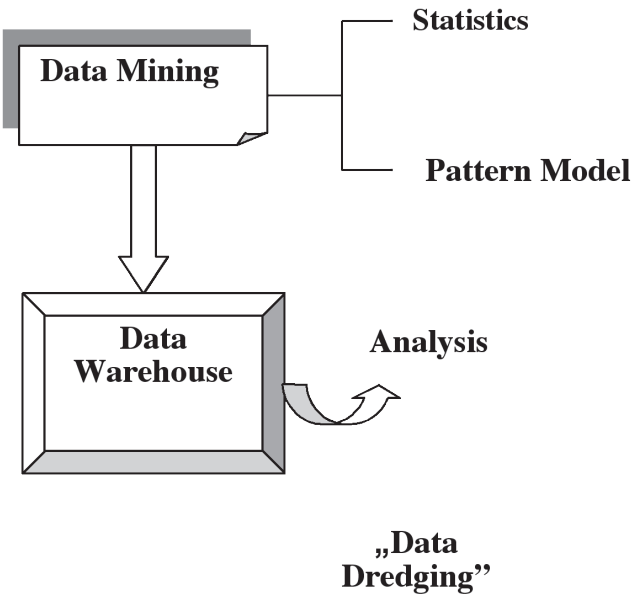


Figure 1 : Elements of data mining

In **data mining** there was made great effort to develop a model of fine granularity and as detailed as possible for data masses. In *"Data Mining for Very Busy People"*, the researchers from West Virginia University and University of British Columbia examined a method that involves minimal differences among the elements of a given amount of data, with the aim of developing simple models representing relevant data.

In essence, **data mining** gives information that would not be available otherwise. In order to be useful they must be correctly interpreted. When collected data involve other persons there occur more problems related to confidentiality, intimacy, legality and ethics.

Data mining consists of a multitude of techniques in continuous development that can be used in order to extract valuable information and knowledge from large volumes of data. Researches into data mining and related instruments would put the accent on commercial applications. Few researches made reference to scientific and satellite data. Despite the conferences' debates on various aspects related to data mining and scientific **data mining**, an exchange of ideas focused on **scientific data mining has not really occurred between scientists and specialists in data mining.**

Data mining is a stage in the process of knowledge discovery consisting in applying data analysis and discovery algorithms that, within reasonable limits of calculation efficiency, produces a particular enumeration of data patterns. It is to mention that the methods, the number of variables considered may be reduced or there can be found invariant representations for data.

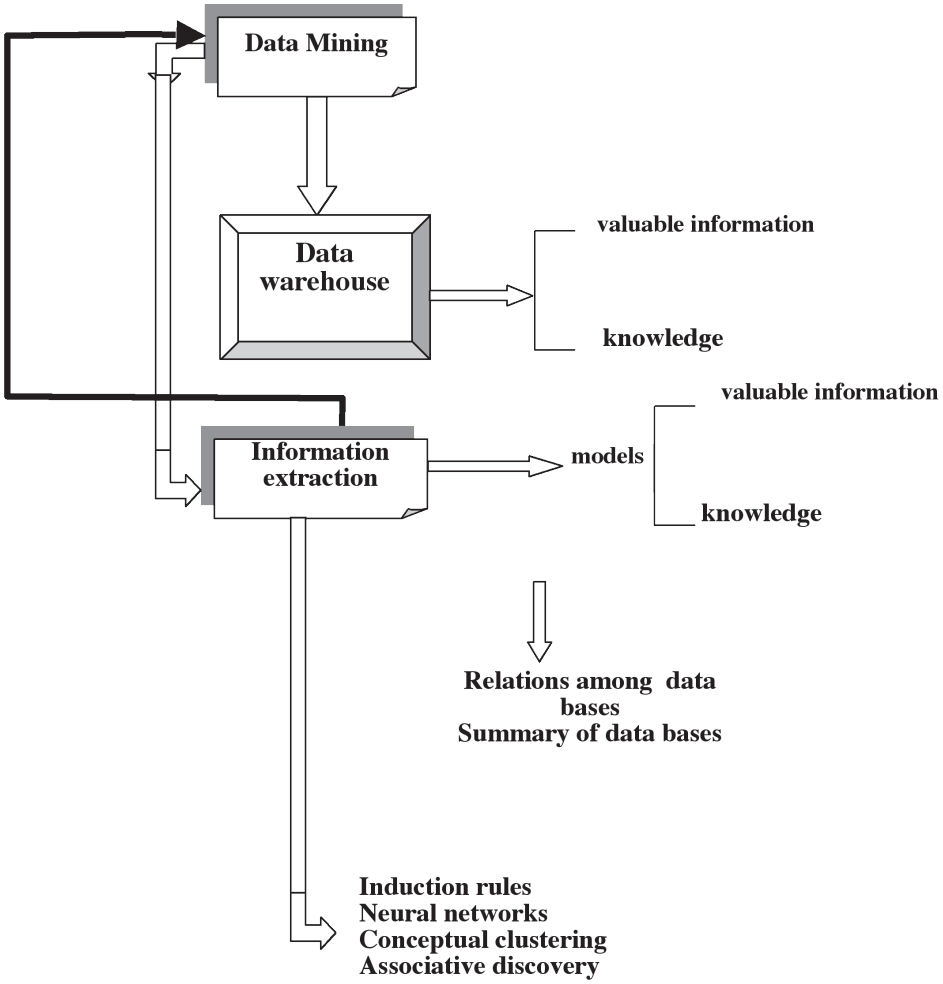


Figure 2 : Data Mining elements and links

In compliance with the Charta of data management solutions provided by IBM, data mining is the process of extracting valid, previously unknown, and intelligible information from a large data base, information used to make decisive decisions in the business field. Information extraction may be used to shape classification or prediction models, to identify relations among the data bases records or to provide a summary of data bases that are mined. Data mining consists of a certain number of operations, each of them being supported by a variety of techniques such as induction rules, neuronal networks, conceptual clustering, associative analysis, etc. In many fields from the real world such as marketing analysis, financial analysis, fraud detecting etc., the information extracted asks for the cooperative usage of more data mining operations and techniques.

Consequently, data mining is an area undertaking continuous development in the field of computers science, which will provide a new and efficient level of information and knowledge discovery that all users in the field of computerized data storage will benefit from.

Data Mining, a new era in computing

Knowledge exploring

"… numerical information is easily acquired and cheaply stored. But what could be done with such a great amount of data?"

In order to make decisions, as revealed above, human beings need knowledge. Therefore, there emerges the issue of data analysis and knowledge extraction from them.

If data are relatively few in number, could either specialists in various fields or statisticians easily and manually fulfill the analysis? Some authors call them "miners" or "manual excavators of data". What are we going to do if we have great amounts of data? How large could a database be in an application? An answer could be found in the Case study box: University Library.

A new discipline was born under several denominations, Data Mining, Knowledge Discovery (KD), Knowledge Discovery in Data bases (KDD), Information Discovery (ID), Information Archeology (IA) etc. Each denomination

may be justified in its own way. Yet, we have to underline that some authors trace differences among them, for example between DM and KDD or KD.

Therefore, what is DM? From the great amount of definitions we choose only two that seem more suggestive and simple. "DM is extracting predictable information from large data bases" or "torturing the data until they confess".

DM and KDD components

The main function of DM is, therefore, to extract patterns of knowledge from data. For this very reason, DM makes use of a variety of statistics algorithms, forms recognition, classification, fuzzy logic, machine learning, genetic algorithms, neuronal networks, data visualization, etc. The variety of algorithms may be grouped in the main components of DM. The number of these components differs from one author to another. Therefore, some authors consider that DM has 3 components, others consider that DM 4 etc. We consider that the main components of the DM are:

- *model* – represented, like any computing model, through a function in a one-dimensional or multi-dimensional space (an assembly of functions), depending on parameters. It can be represented as either a linear function of parameters or a probability function or fuzzy function etc. The model is conceived with several algorithms, such as classification and clustering algorithms;
- *preference criteria* –can be of different nature, some of them being based on ordering, others on interpolation or on the best approximation;
- *selection algorithms* – leading to a selection of three important elements occurring in DM, that is: the model, which is selected from the models basis, data, which are selected from the database and represent parameters, and the criterion or criteria of preferences, which is selected from the criteria base;
- *setting the deviations* – generally consisting of algorithms determining the deviation or the stability; a specific category of such algorithms are the statistical ones, through which there are established the model's deviations as compared to the ideal one.

The authors that make the difference between DM and KDD consider KDD as being an iterative and interactive process that includes DM. Therefore, as part of KDD the knowledge extraction is accomplished by following the steps:

- *learning the application-related field* – consisting in the acquisition of knowledge about the initial condition as well as about the particular goal of the application;
- *establishing the data set* – to which is applied the process of knowledge extraction; most authors draw the attention upon this stage called the preprocessing stage;
- *cleaning and selecting data* – which is a complex process, specific to data warehouses, coping with noise elimination, dimension diminishing, data replacing, etc;
- *applying the DM procedures* – which is in fact the most important step of KDD process;
- *interpreting results* from user's point of view – which is a decision-making stage; if he is not satisfied with the result, he can go back to any of the previous stages;
- *using the knowledge discovered* – which is the final stage; this use is accomplished by incorporating them within integrated knowledge systems or by concluding simple reports to those interested.

It is obvious that the specialists considering KD or KDD synonyms with DM, state that these stages are part of DM stages.

Link between DM and other fields of computing

DM and KDD are connected to new fields of computing. Our intention is not to present all these connections. However, we underline that the closest fields to DM and KDD are OLAP (On Line Analytic Processing) and DSS (Decision Support Systems). There are many presentations, more or less complete, on OLAP and DSS. According to them, OLAP is a way to use data warehouses, usage that supposes, on the one hand, on line access (OLTP – On Line Transactional Processing) and, on the other hand, a multidimensional (vector-based) analysis of large data bases. DSS is an assembly made up of data bases and data warehouses, as well as other assemblies of useful information, together with soft products conceived to draw up reports, data analyses and to implement optimization algorithms in view of supporting the process of decision-making undertaken by business people. Since these two fields are strongly connected to DM they determine a series of OLAP or DSS providers to sustain the DV delivery. "Giving a weak interpretation to the notion of Data Mining, the OLAP providers may say they deal with DM" states Steve Smith, director of advanced analyses at Pilot Software. "One of the

lines of demarcation between the two fields is the automation degree. The second one is the degree of using the statistical means to determine what is interesting and relevant".

These are definitely in favor of DSS. OLAP and DSS use fewer algorithms and depend on the requests and hypothesis delivered by the user. As far as the answers are concerned, the difference between OLAP and DSS, on the one hand, and between DM and KDD, on the other hand, may be associated with the difference between the answers given by a data knowledge and bases. Consequently, for instance, a typical way to formulate a question in OLAP and DSS is: "*Did users from Norway access more about vehicles as compared to the users from Romania last year?*". The OLAP and DSS answer takes the following form: "*The Norwegians accessed 12 000 publications about cars, whereas, the Romanians have accessed 10 000 within the same period*". This answer is therefore clear but limited to hypothesis and, therefore, rigid. The answers in DM and KDD are much more flexible. In their case, the questions take the following form: "*Present a model that identifies the most predictable characteristic of the population that is to access publications about cars*". Learning from past experience, DM will answer: "*It depends on period and department. Therefore, for example, last year users from Norway accessed more publications about cars (12 000) than those from Romania (10 000)*". The advantage is that DM and KDD may discover more sophisticated and subtle answers that OLAP and DSS cannot detect.

There are even more differences, among which one of the most important is the reasoning way.

Therefore, the reasoning in OLAP and DSS systems is deductive, while in DM and KDD it is inductive.

On the other hand, it must be underlined that the differences between DM, OLAP and DSS have the tendency to disappear. The DM algorithms and facilities have been more and more often introduced in DSS and OLAP products. For example, AT & T, which is a great OLAP user and provider, has started to work on a new product DM, named Nearest Neighbor System that classifies the users according to similarities. At their turn, the DSS providers have specific visualization systems of data bases; therefore, introducing some DM functionalities is a normal option. It is not to wonder that the OLAP and DSS providers, such as Pilot and Comshare have rapidly oriented towards DM.

One of the favorite fields of DM usage, by OLAP and DSS and producers is Internet enhancement. This fact appears as natural, viewing the heterogeneous character of data bases and warehouses in this case. Therefore, for instance, Tree Software has introduced a Web gateway at a price of 10.000 dollars, a gateway that is a site between the Web server and a multidimensional OLAP database of type Essbase of the company. The product assures the writing and reading of

business information analysis furnished by the browser Web, which will facilitate its spread on DSS, OLAP and DM markets.

Data Mining applications in digital library – Bibliomining

Origins and definition of bibliomining – Theoretical Concepts

Bibliomining derives from the terms "bibliometry" and "data mining", since the aim is to benefit from social networks that justify the functioning of both bibliometry and user-oriented data mining, through the intermediary of a unique center of data storage. Why should a new term be created for data mining in libraries? The concept is not new; data mining has been occasionally analyzed within library science meetings beginning with the last half of the 1990 [5]. The challenge resides in the terminology used; the packages of data mining comprise a library of various algorithms. That is why searching for articles in a particular field may lead to documents that are not on topic, such as "Re-usage patterns aiming at data mining in Library for the applications selected by user" [108]. In view of facilitating the activity of specialists in libraries-oriented data mining and in view of avoiding the confusions of library-oriented software for data mining, Nicholson created the term of "bibliomining" for the paper from 2003 by Nicholson and Stanton.

In order to better conceptualize the bibliomining concept it is useful to first conceptualize the data required for traditional bibliometry and user-oriented data mining and then to see how they can be combined to create the bibliomining concept.

Bibliometry

Bibliometry is based on quantitative exploitation of document-oriented scholastic communication [9]. Figure 3 presents a part of the data used in bibliometric research and the connections among different papers. The papers have authors and collections (periodicals, publishing houses, libraries) associated to them, and the papers are interconnected by the intermediary of references, work paternity, common terms or other aspects of the conception and publishing process.

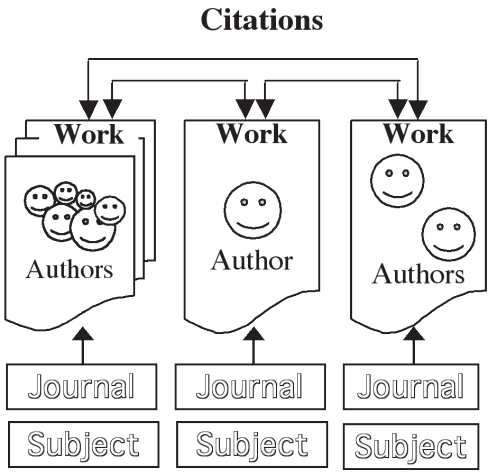


Figure 3: Connections among the papers used in bibliometric research

The traditional bibliometric data imply information regarding the paper's conception, such as paternity and cited works. In addition, the metadata associated to the paper, such as the general thematic or particular periodical where it occurred are connected to data viewing the paper's conception. The association of these data enables the researcher to understand the context in which the paper was conceived, the impact of citation on long term and the differences among fields as regards their patterns of scholastic results.

The analyses carried out in traditional bibliometry were based on frequency; still, many recent bibliometric studies resort to visualization and data mining in order to explore the patterns by creating these materials [90; 11]. A part of the concepts explored includes the paternity frequency in a filed, the generality degree of words used and the discovery of a central set of frequently cited works [10]. The reference integration among works allows a rich exploitation of relations between scientists and subjects tackled; the connections among papers are used for automated searching of information and erudition visualization [155] and of social networks [136] for those involved in the creation process. Many recent bibliometric applications involve Web-based resources and hyperlinks that emphasize or substitute traditional information linked by references [9].

User-oriented data mining

A popular area of data mining not only in the commercial sector but also in the scholastic literature consists in investigating the way users explore Web spaces. These studies are centered on Web pages accession by a certain user (or IP

address). There are discovered usage patterns through data mining that are used to personalize information offered to the user or to improve the information service [156; 38]. Figure 4 reveals a part of the data used in this user-oriented data mining. The aim of this figure is to demonstrate that in user-oriented data mining the links among papers come from a common usage. If, for example, a user accesses two papers during the same session, when another user visualizes one of these papers, the other might as well arise his interest. This figure presents the links among works that result from users.

A record in this data source represents a unique accession of the data resource, and the metadata attached to this recording include any available

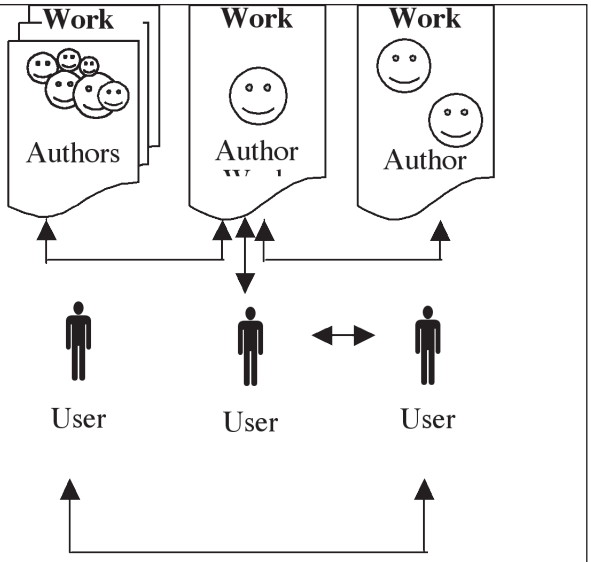


Figure 4: Data for User-Based Data Mining

identification as regards the user, information related to time and datum and data related to the corresponding Web site. Certain studies append metadata with regard to the work accessed in view of improving the recommendation algorithms [51; 87]. The patterns are oriented towards the understanding of the manner in which the users explore the informational space; if there is any way to identify a user between sessions (through a cookie or a login), the users' behavior may be detected throughout time. Since many digital library services require a certain type of login to access materials bought or rented, this type of data mining in use is possible and will be useful in the decision making process.

The challenge of implementing this type of exploitation in developing the activity of a digital library consists in that of user intimacy. The private character

of information making reference to the user that can be identified as person represent a real preoccupation during the bibliomining process.

A commonly inferred solution consists in coding the user's identification data in the center of data storage. This would allow a detection of articles within a period of time and it would not allow an immediate identification of users in the center of data storage. Yet, the substitution of the user's ID with a code is not appropriate as anyone wishing to find out information regarding the user's behavior may resort to the codification diagram to find the codified ID for a user and to find the information regarding the user. That is why the codification may represent a tempting choice but if the codification procedure is reversible or a codification study is maintained, the user's private character still represents a worry.

Data storage center for bibliomining

Both bibliometrics and bibliomining for Web usage have a data field in common – the work accessed. On the one hand, the bibliometrics is focused on work's creation, and, on the other hand, the exploitation for Web usage is focused on work's accession. The combination of these two sources of data in a single center of data storage allows the researchers and library managers to fully understand the information space created. Figure 5 presents the model for data resources in bibliomining, which demonstrates the concept of connections among works on the basis of both the creation process and users' population.

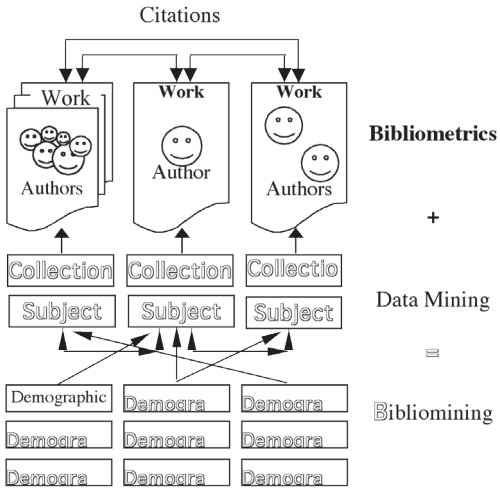


Figure 5: Data for Biblio-Mining

The bibliomining is defined as combination between data mining, bibliometrics, statistics and reporting instruments used to extract artifacts patterns based on behavior in library systems [113]. It was rarely implemented in full form, as it was presented here, due to the complexity of problems related to data. By presenting the model the goal is for the institutions developing centers of data storage for digital libraries to keep in mind the complete center of data storage in bibliomining as aim while developing smaller projects. The integration of bibliomining within current research and evaluation will allow managers and researchers in the library to make a more complete idea upon the resources held by their library and on the way they are accessed by users.

A framework for data

The data that can support the links between both creation/publication and usage have to be found in the same center of data storage in view of allowing the whole bibliomining process to develop. A conceptual frame for these types of data is useful to determine the fields that are better to be excluded from the operational systems. There are three parts making up this frame - data regarding the work, data regarding the user and data regarding the service. These three parts will be interwoven to represent a usage that is at the basic level for the center of data storage. The case of using a library service connects one work (or works), a service and a user in the center of data storage for bibliomining.

I. The first section of the center of data storage comes from **work**. This will comprise three types of fields – fields that were extracted from the work (like title or author), fields that were created as related to the work (like subject heading) and fields indicating the format and work's location (like URL or collection). These items of information may come from a MARC recording, Dublin Core information or from the management system of library content. This area may be converted into bibliometric information, such as references or connections with other works. This may ask for the extraction from the original source (in case of digital references) or connection with a database for references. A challenge of creating this link consists in the fact that providers currently report a use of electronic resources at the title level, while many bibliometric explorations begin at the level of article. There are required standards for reporting at the level of article linked to COUNTER formats; once created, the providers may provide more information regarding the specific elements used at their sites.

II. The second area of the center of data storage comprises items of information regarding **the user**. As previously mentioned, this is the place where the

demographic surrogate is to be stored. Supplementary, there will be stored other fields that come from user interferences. For instance, the user's IP address may activate usage place interferences. In the case of university libraries the IP address could be used to deduce if someone comes from off-campus, on-campus or from the library. In certain cases the IP address could provide information regarding the building or the computing laboratory where the request comes from. A similar inference from the public library may come from the zip code. Both inferential connections may provide demographic keys with regard to the users' groups, but neither will provide a complete demographic superposition. However, at an ample set of data there may be deduced patterns from these best demographic keys.

III. The third area comes from **the main reasoning of library utility** – to connect users to information, especially by the intermediary of **works**. The most difficult part to be conceptualized is the library service since there are many and various types of services. **Searching, circulation, references, inter-library loan** and other library-oriented services have in common fields that may be captured in the center of data storage. Supplementary, each of them presents a set of appropriate fields for each type of service. A center of data storage adequately conceived may handle both types of data; this situation allows the evaluation of a certain type of service or the acknowledgment of library utility by the intermediary of its multiple services. The fields common to most services include time and data, library staff involved, location, method and if the service undertaking was made in association with other services.

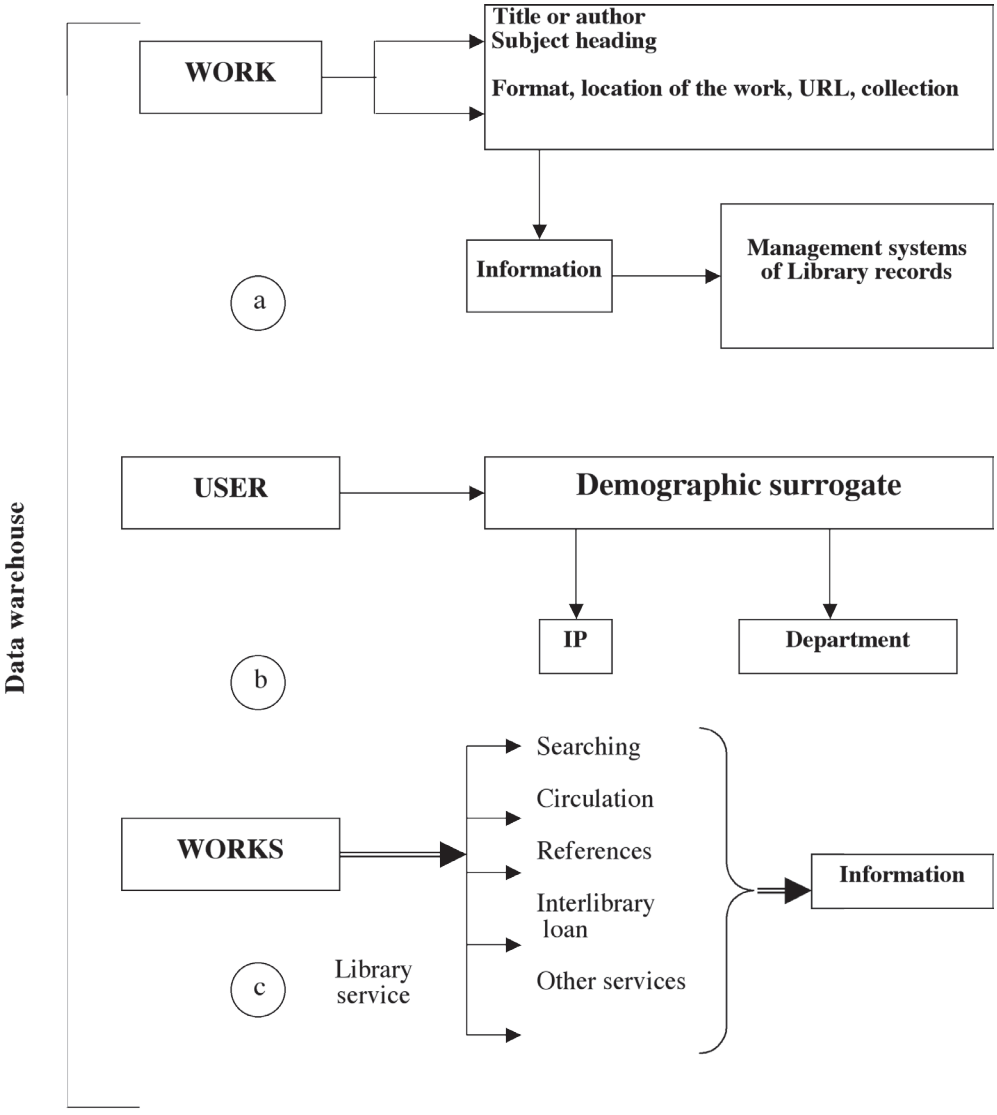


Figure 6

Each library service presents a set of adequate fields. For instance, the searching presents the searching content and the following steps to be undertaken. The inter-library loan presents the cost, a provider and an interval to fulfill the request; the circulation presents information with regard to the acquisition of work and duration of circulation. As regards most of the decisions, the necessity to make all the decisions assigned to the library and the necessity for scientists'

research should guide the captured fields while still keeping the user's intimacy. In order to support the exploitation, Figure 7 comprises many other components and fields that demonstrate the conceptual frame for the storage center of data within bibliomining.

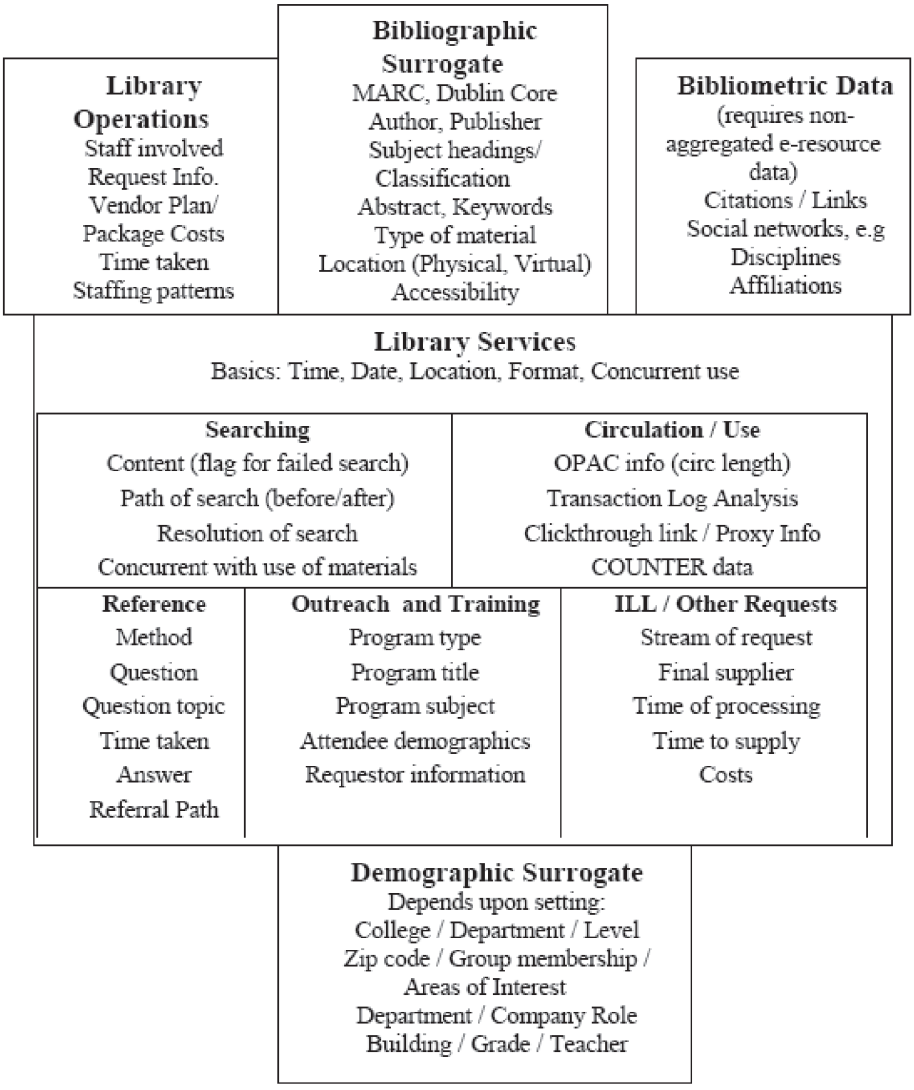


Figure 7: Conceptual framework for date types from data warehouse

Using the center of data storage for bibliomining

Once the source of data is collected, it allows not only the creation of traditional reports and measures but also the creation of new exploitation opportunities. This section presents various instruments that are useful in exploiting the center of data storage; the goal of this section does not consist in explaining the way to use each instrument but in presenting the instruments within the context of structure of the center of data storage for bibliomining and the way in which these instruments could be further developed and implemented.

Traditional reporting

Traditionally, the persons responsible for making decisions in a library investigate the media in order to understand the way their service is operating. These measures may be all created together with this center of data storage but, at the same time, they present the advantage of enabling the managers and persons in charge of making decisions to ask other questions. If there are collected amounts of data exclusively and the subjacent data are eliminated (or are never put at disposal), then the ability to ask new questions is reduced to a great extent. The questions must be formulated in time and once the reports are drawn up they are difficult to modify as it is likely that the measures are not comparable in time. An exclusive argument to keep the amounts of data consists in the fact that it protects users' intimacy; despite all these, the utility of demographic surrogates allows the library to keep supplementary data with regard to its users and services while protecting their intimacy.

The advantage for the center of data storage consists in the fact that new questions may be formulated not only with regard to the current situation but also as far as the past is concerned. Since the center of data storage represents a collection of past actions, from past actions there may be collected new media and amounts of data. This allows the persons in charge of evaluation and measurement to think up new questions and then to look through these reports in view of understanding the tendencies. It would be an impossible task if only the amounts of data were kept; there could be formulated exclusively new questions and there would not exist any background to implement measures in perspective.

Furthermore, libraries may understand more easily the behavior of different demographic groups within the library. The amounts cover the subjacent patterns that might be emphasized if the same measures would be collected for different groups. Since the library addresses sets of population, there will be situations when each group resorts to library for different informational necessities and by following different paths. As simplified example, let us suppose that a group of

population uses the references and another group uses the electronic resources; in this case, the fusion of all library users will have unsatisfactory and unusable results – half of the time is taken to resort to references and the other half is taken to resort to electronic resources. Considering these groups from different perspectives, these differences will be outlined, which will lead to much more practical results.

Data mining

The goal of data mining is to explore the set of data for new and useful patterns. Data mining may have a definite goal and, therefore, we have in mind a certain goal or a certain topic area; data mining may not have a definite goal when the goal is simply to discover something interesting. Data mining includes some of the techniques that have been already discussed, as well as other instruments coming from statistics and artificial intelligence, such as neuronal networks, regression, clustering, rules and classification. Data mining represents the process of collecting a purified set of data, of generating new variables starting from the existing ones (such as the creation of a flag yes/no from a numerical variable), data sharing in sets model edification and test sets, application of techniques to sets of model edification in order to discover patterns, use of test sets to assure that the patterns are more generalizing and then conformation of these patterns by someone that knows the field [7]. These patterns will constitute then seeds for more detailed explorations that may result in new explorations, in new reports and even in new cumulative measures that may become integrant part of the measures collected on a regular basis.

In view of fulfilling data mining there are required non-cumulated data. Even if a library does not hold instruments or expertise necessary to explore own data it can assign this task to researchers if it maintains a non-cumulated version of data from their systems. By using the concept of demographic surrogate, the library will protect the identity of its users, enabling at the same time other persons to discover attached patterns from demographic data. While substituting individual users with demographic surrogates loses a part of the information, it is implied the hypothesis that the patterns associated to demographic groups may be still discovered. These patterns based on demographic information are the patterns that prove to be useful in allowing librarians to better adapt their services for groups of users. It is important to involve the library; in view of creating useful patterns the librarians have to get involved in naming the taxonomy for demographic surrogates.

A sub-area of data mining that proved to be useful in the process of digital library evaluation is the Web-based data mining. This branch of data

mining begins through a transactional log in a Web server that allows digital library services to detect the information selected by users. The challenge when investigating the exploitation through a transactional log is data purification, since the transactional logs are difficult to use. This concept was explored by Srivasta et al. [143] and was applied directly to digital libraries by Bollen et al. [8].

Data mining represents an area interwoven with bibliomining; yet, it is not directly supported by the center of data storage for bibliomining. In text mining, it is the works' context that is to be explored. The center of data storage for bibliomining, as defined herein, would not support text mining due to the fact that works are represented through demographic surrogates. Despite all these, if one wishes to carry out a project of text mining one would have to take all textual data and append fields from the center of data storage for bibliomining. This could substantially enhance the projects oriented towards text mining, as it would explore items of information comprised within works, metadata making reference to works and information on work's users.

Kostoff et al. [90] – found out that the combination between text exploitation and bibliometrics allowed a better comprehension of the community of authors than each of them used separately. They began by four works and extracted all the works they quoted. Then they resorted to text exploitation in order to search for sets of themes in the text of this corpus of papers. A key discovery consisted in the fact that the papers which quoted the original works, arising however from a different discipline, could be removed from the corpus of works that identified themselves with the adequate discipline. The disadvantage of this process consisted in the quantity of time dedicated to the manual processing required for analyzing only the four original works and quotations of first generation. Nonetheless, the work demonstrates the potential of the interweaving of concepts from bibliometrics and data exploitation for improving the comprehension of a certain situation.

Conceptual Framework for the Specialists in Biblioteconomy and Information Science

Unlike the decision-making persons who need to focus upon a certain library system and would obtain the greatest benefits from a holistic understanding of that certain system, the experts in biblioteconomy aim at amplifying the understanding of biblioteconomy upon a larger scale by an activity directed towards generalisable assertions. With a view to comprehending the context for bibliomining within this research, we fix our eyes upon archaeology. Archaeology

constitutes an adequate model as the archaeologists' process of collecting artifacts in order to make assertions with respect to the humans living in a certain area is related to the process of the data warehouse creation for bibliomining. When the users wander through a digital library, they leave behind artifacts based on data of their visit. These artifacts may constitute a record for a Web log, a Proxy server login or an interrogation point used in a research within the digital library. During the phase of data warehousing, these artifacts are collected from different locations associated to the library system and are reassembled with a view to obtaining a more complete comprehension of the users who "lived" in that digital space. This concept is developed as "Internet Technology" by Nicholson [117].

In many library studies, these artifacts of the use are collected, cleaned, numbered and displayed by presentations type conference and articles, with an indication to the way in which these statistics are compared to those of other libraries. This situation is similar to archaeology until 1960: researchers focused upon a site, collected artifacts, exhibited them to others and then endeavored to operate connections between the artifacts from this and other expeditions. However, during these last 50 years, archaeologists have extended their conceptual frame by means of the "new" archaeology and of the post-processing archaeology. The new archaeology focuses upon the increase of the knowledge base, instead of simply collecting artifacts during the searches [85]. This fact may be related to the current state of digital library evaluation; rather than formulating specific questions which should advance knowledge with respect to the users of the digital library, many researchers resort to one virtual "dig" after another, collecting more and more measures with no constructive advance towards the knowledge [117]. Post-processing archeology enhanced the importance of the awareness that the situation was much more complex than the artifacts stated; issues such as the social context and the community influences had to be taken into consideration [85].

The resulting framework of the research is much more complex but it may result in creating much more generalizing knowledge assertions. Archaeologists begin by examining the artifacts available for patterns. They create generalizations from these patterns and develop hypotheses. Subsequently there are created studies for every hypothesis, and the archaeologists revisit the site with a sample methodology, collect new artifacts, talk to the people in the field and test the hypotheses. Eventually, these hypotheses are tested in other organizational frameworks for a subsequent scientific knowledge. This path is known as the hypothetical-deductive-inductive method [142].

1. Induction
(Recognition of the pattern by bibliomining)

2. Deduction
(Logical analysis of the patterns in order to obtain generalities)

3. Prediction
(Creation of the hypothesis from the generalities)

4. Testing
(Developed research in order to test the hypotheses)

Target: Knowledge about the use of the library

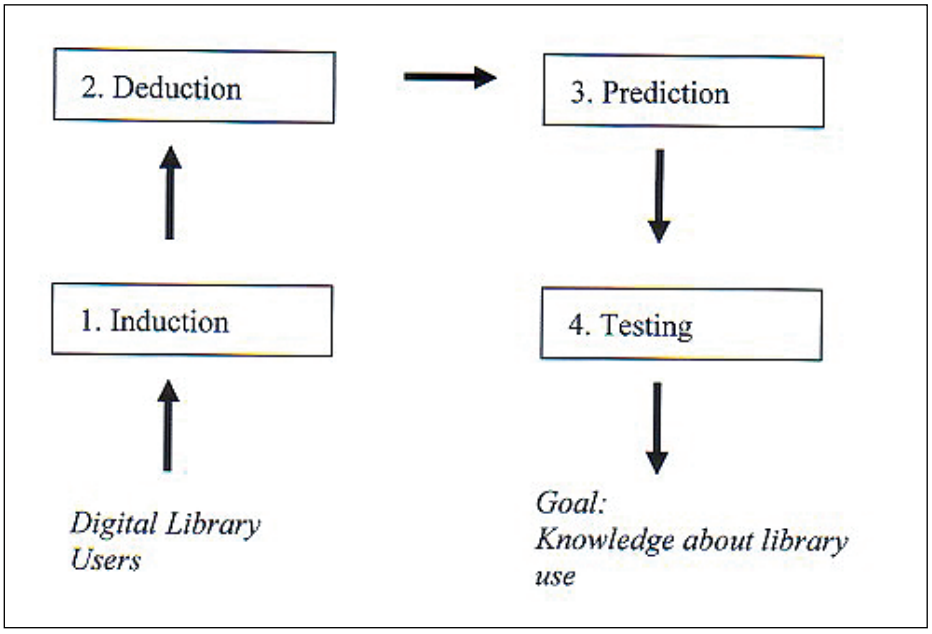


Figure 8. Framework inspired from archaeology for the research within the digital library

Figure 8 presents an application of this framework to the research effected in the digital library and it places bibliomining in context with the rest of the research process. Similarly to the previous case, bibliomining stands for an important step within the process, nonetheless representing but a simple step in a more complex process. First of all, there are collected the artifacts of the digital library,

and bibliomining is used in order to determine the patterns. These patterns may subsequently lead to generalities, which may inspire hypotheses. Subsequently there are designated studies which should test these hypotheses; these studies may be quantitative or qualitative (or both), and they may imply addressing to the users of the digital library or collecting new data. The results emerging from these studies may lead researchers on to the path of new knowledge.

Post-modernist Archaeology

Post-modernist archaeology avoids the process which produces generalizations about a culture and focuses instead upon the individuals of a certain culture [117]. Any culture is made up of individuals, each of them making their own decisions in life. In order to comprehend a culture, a researcher has to understand the individuals who make up that certain culture.

Due to the accent laid upon the users by the researchers in information science, such as Dervin and Nilan, important research have been advanced with respect to the importance of the users for evaluations based on artifacts during the evaluation process within libraries. Saracevic and Kantor argued the importance of the users' implication in learning the relevance and in the usefulness of the resources retrieved within a system. The holistic framework for measurements presented in Table 1 [115] presents the relation between bibliomining or the internal view of the usefulness of the system by artifacts in guise of data, and other measurement areas, which include:

- Internal view of the system, focusing upon the information and services offered;
- External view of the system, implying the users in order to find out whether the information offered meet the requirements of a certain user; and
- External view of the use, implying the users in order to find out whether the information proved useful to the user.

	Subject	
Perspective	System	Use
Internal (System of the library)	Standards of procedure	Interactions registered with the interface and materials, Bibliomining
External (User)	Approximate use	State of the knowledge Value of the works

Table 3: Matrix of measurement extracted from Nicholson's measurement frame

An advantage in examining the information consists in the implication of the users of a system in the evaluation process. We may speak of certain scenarios for the users' implication within the evaluation:

- In case of a log-on procedure, the users may comment how satisfying the results proved for their search and how useful the information was in meeting their requirements. In order to collect these viewpoints, an opinion poll effected after the interaction with the system will prevent the relevance decisions from modifying in time.
- In case of no use of a log-on procedure, the construction in the deduction methods during the information delivery process may contribute to the capture of the relevance decisions as for a piece of information, by a user. The necessary data may be captured by requesting the e-mail address and by continuing with an opinion poll about the utility of the information.
- Another approach consists in localizing the search themes for the typical users of the system. These users may operate their search according to their own needs or they may be offered searches, the measurements being subsequently collected by this process. This is the least desired scenario and it may be the only one available.

While bibliomining may contribute to the comprehension of the existent use patterns, it can't help evaluators to understand the documents deemed relevant by people (in comparison with the documents they used). Moreover, the results in bibliomining are restricted by the system resorted to; if a characteristic is not offered by a system, no information in connection with that specific characteristic can be obtained. This is the reason why bibliomining, the same way as archaeology can't offer the truth, as the truth about the search of information is in the possession of the person effecting the search. Bibliomining may retrieve information about the interactions between an individual and the system and it can retrieve nothing with respect to the user's mental state. Consequently, the collection of these measures allows the researcher to bring qualitative elements in the quantitative bibliomining process.

Completion of the Cycle

The addition of post-modernist archaeology to the Hypothetical-Deductive-Inductive cycle allows researchers to place themselves behind the data in the system in order to better comprehend the users. The processes have been separated from the products in South's cycle in order to better integrate the requirements of a digital library. Making use of the entire cycle, the researchers may operate the passage from the description to the improvement of the information base about the users of the library. The significant processes of this cycle are:

- Collection: gathering the artifacts about the services of the library, about the users and resources,
- Induction: recognition of the patterns by bibliomining and visualization,,
- Deduction: logical analysis of the patterns in order to produce generalizations,
- Prediction: creation of hypotheses out of the generalities, and,
- Testing: research developed in order to test the hypotheses which imply both data and users.

It is worth knowing that we can never know the user's experience only from the artifacts within the system; not even a dialogue with user after the process offers a valid description of the user's way of thinking. The same way an archaeologist does not know the truth about the people who developed in a certain culture, we neither can know the truth about our users only looking at the artifacts. We have to work directly with the users in order to test our hypotheses with respect to their behavior.

This resulting archaeological frame is an application of the traditional scientific process. Librarians and other specialists who offer web information resources do not regularly use research projects based on hypotheses when their services are evaluated. The result is that the science of biblioteconomy and information did not follow the same development rhythm of other sciences. The advantage of the archaeological approach is that it may be of help for the non-initiated in the scientific method to conceptualize the way in which this method applies to their own information service.

Conclusion: Passage from Evaluation to Comprehension

The final and longest research bibliomining can inspire consists in the improvement of the comprehension of digital libraries at a generalized, perhaps even conceptual level. These data warehouse centers will combine the resources hitherto not traditionally available for researchers under this combined form. As in the archaeological frame presented above, this is the point in which bibliomining may inspire new questions for the research process. For example, what connections may be effected between the demographic information about the user and the authors' social networks based on bibliometrics? How extensive is the influence of the written and quoted works by an institution upon the use patterns of the library services by the students? How do the use patterns differ between departments and demographic groups and what measures can the library adopt in order to personalize and improve the existing services? They may imply qualitative measures in order to support the quantitative data, or they may include the collection of other types of data from other measurement

fields which should enhance the bibliomining data. The exploration of these new questions vouched for study may lead to a thorough comprehension of this science of biblioteconomy.

Moreover, these vast warehouses of multi-system data may allow visualizations of a knowledge space by means of the comprehension of the connections among the resources resorted to or among the common points from the references transactions. Taking over the methods currently of use in the visualization of the bibliometric data, expanding them in order to include the connections among the papers and adding animation with a view to demonstrating the way in which these data modify in time may allow the understanding of the evolution of the knowledge spaces.

Due to the nature based upon artifacts of the digital library services, the decision-makers and the researchers can understand the users' information search behavior, previously available only within a restricted control environment of the research. These artifacts based upon data may be collected, deprived of identity and combined from different systems and services into data warehouses which offer thoroughness to the decision-makers and information ampleness for the users, necessary for a better understanding of the digital libraries. Resorting to bibliomining in guise of pivot of the conceptual frames allows the users and the decision-makers to cooperate for the development and solving of the research issues both of short-term interest and of long term impact.

Stages of a Bibliomining Project

Bibliomining, which means data mining in the framework of a library, stands for a fragment of a more comprehensive project. The process as a whole is known as Information/Knowledge Discovery within the Data Bases (KDD). This page will enhance the way in which KDD may be made use of in the framework of the library environment.

1. Identification of the theme. The first step consists in the identification of the theme the library subject deals with. There are two types of bibliomining – predictive and descriptive. Predictive bibliomining may either predict a future event on the basis of the past and present state of affairs, or predict a current event which is difficult to visualize on the basis of a small group or of measurements carried out in the past. Descriptive bibliomining aims at describing a current situation.

2. Creation of the Data Warehouse. The data sources which may be of help in the field of the chosen theme must be identified. After this stage, the data must be collected from the adequate systems and combined within a

unique data warehouse. Moreover, the data are purified and there is carried out an operation upon the default values. This stage may occupy up to 80% of the time assigned to a bibliomining process; however, the final results greatly depend upon the accomplishment of this stage. If a librarian deems bibliomining useful, then he must dedicate a significant temporal lapse in order to create a constantly updated data warehouse, which should extract the data out of the operational system, should purify them and should deposit them upon a well fundamented basis. In the future, this investment will greatly facilitate the development and the accomplishment of the bibliomining processes.

3. Refinement of the data. Subsequently, the data adequate to the bibliomining process you focus upon are taken into consideration. New variables (ratios and classifications) may be generated from the original variables. It is possible for a detection and confrontation of the variables with the extreme values to take place. Within this stage there is necessary for an operation upon the default values ignored within the previous stages to be carried out.

4. Exploration of the data. Within this stage, the real bibliomining process unfolds. On the basis of the expected result and of the types of data, different techniques and reports are made use of in order to discover new and operable patterns (Then a miracle occurs …).

5. Evaluation of the results. The patterns which have been discovered must be intelligible for the librarians who operate upon the field the theme affiliates itself in. If the pattern proves unusual and contravenes to the "common sense", then it probably represents a flaw, en error with respect to the data. If this thing happens, the operator has to check the individual records with a view to discovering what generated the respective pattern. In case a predictive pattern was created, then it may be applied to a resistant sampling in order to test the reliability of the model.

6. Report and implementation. In case predictive patterns were resorted to, then they may be implemented. If necessary, they may be implemented upon a reduced sampling of real world data in order to detect their efficacy before their implementation on a large scale. There may be created and advanced reports to the members of the staff who are involved. Many a time, this may generate more questions, a fact which may send you back to step 1.

IV Data Mining Applied in Automatized Library
- a different approach –

Automated analysis

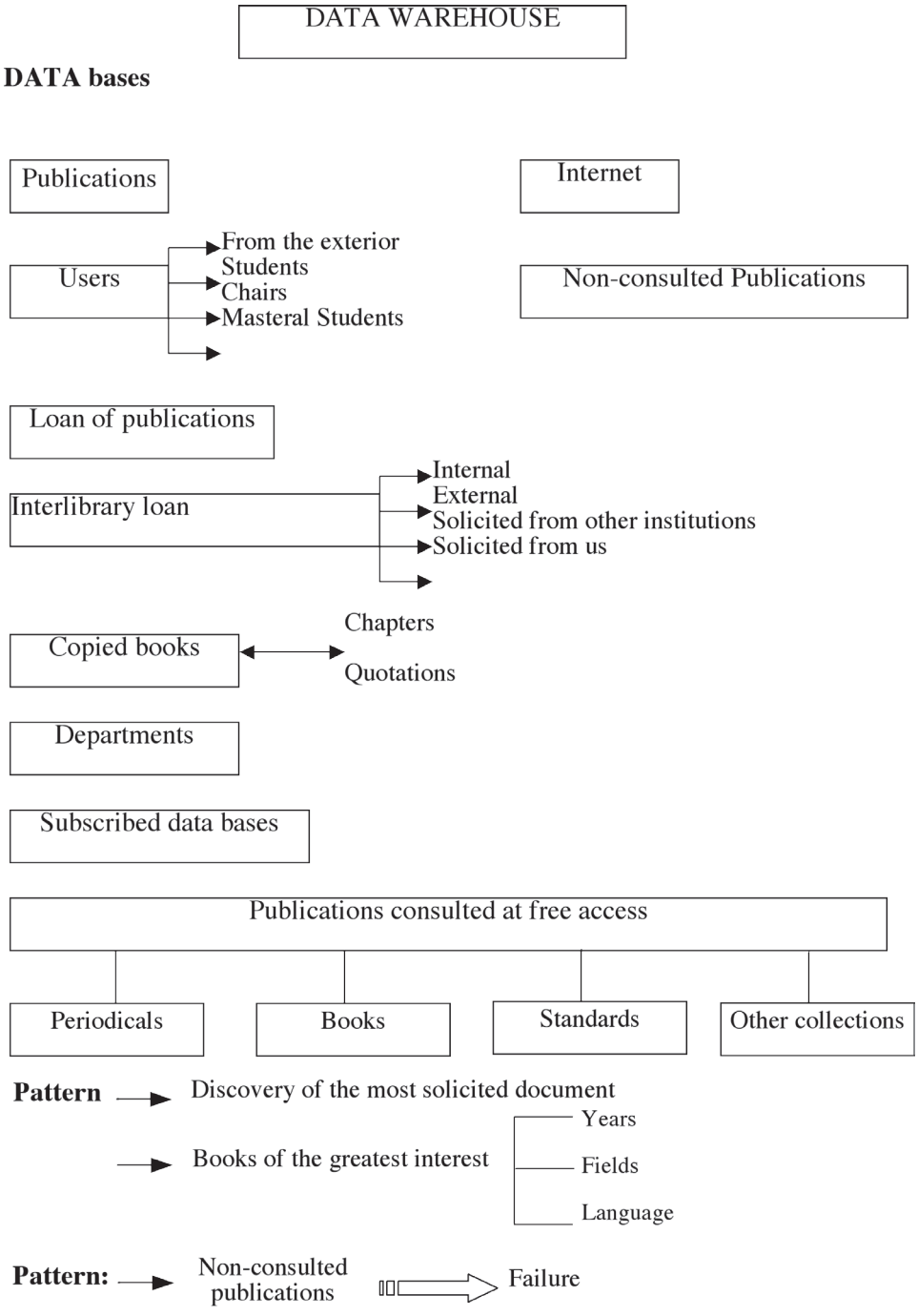
Data mining computerizes the process of thorough research of the past data with a view to discovering new information. This is the most important difference between data mining and statistics, in which case a certain model is divided by a statistician in order to treat a specific analysis problem. The same difference operates a distinction between data mining and expert systems in which case the pattern is built by an engineer out of rules extracted from an expert's experience and work.

The phases of the automated retrieval also operate a distinction from OLAP, simple questions and report tools, which are resorted to when verifying the hypotheses formulated by the user.

Data mining does not base itself upon the user in defining a specific question, only in formulating a target such as the identification of the most accessed works.

Large or Complex Sets of Data

One of the attractions of data mining consists in making possible the analysis of very large sets of data within a reasonable period of time. Data mining is also convenient for complex issues which imply relatively small groups of data, in which there are however many fields or variables to be analyzed. Nonetheless, for simple and relatively minor issues of data analysis, there are simpler, cheaper or more efficient solutions.


DATA WAREHOUSE
DATA bases
Publications
Internet
Users
From the exterior
Students
Chairs
Masteral Students
Non-consulted Publications
Loan of publications
Interlibrary loan
Internal
External
Solicited from other institutions
Solicited from us
Copied books
Chapters
Quotations
Departments
Subscribed data bases
Publications consulted at free access
Periodicals
Books
Standards
Other collections
Pattern
Discovery of the most solicited document
Books of the greatest interest
Years
Fields
Language
Pattern:
Non-consulted publications
Failure

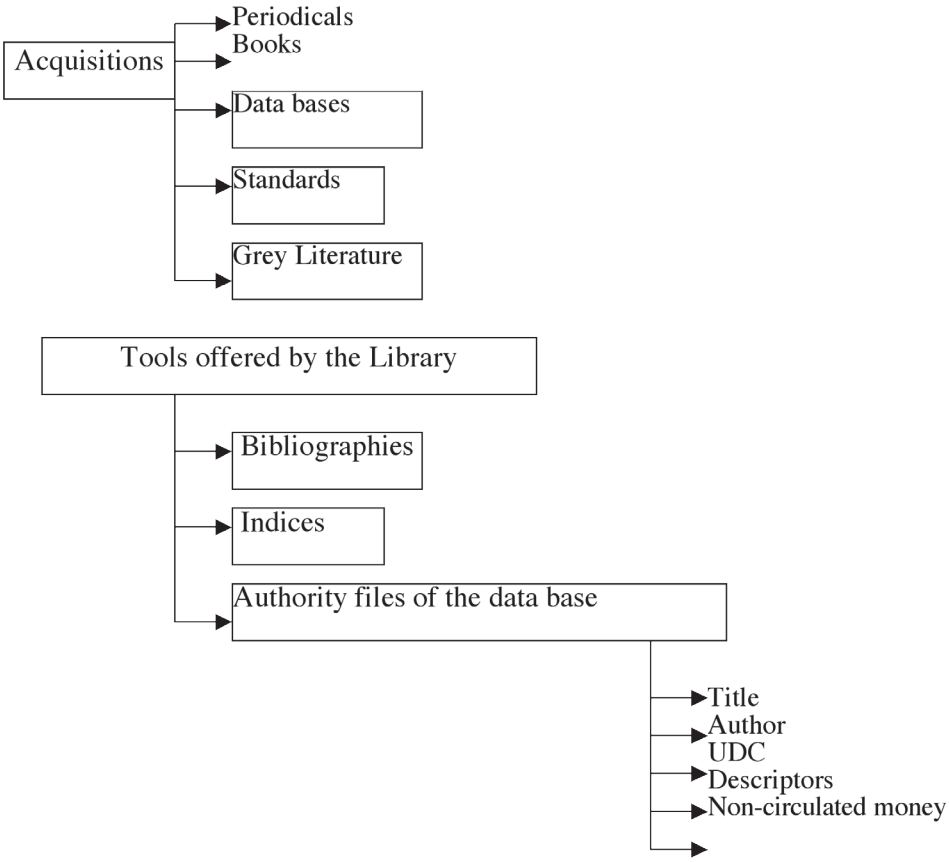


Figure 9: Data Warehouse Components

Discovery of Significant Patterns

The target in "data mining" consists in discovering relations among data which might provide useful meanings.

"Data mining" tools may scan data bases and may identify patterns, previously hidden, in a single step.

- One example, of pattern discovery, consists in the analysis of the most accessed works, fields, authors, publishing houses.
- Other problems of "pattern" discovery include the detection of non-circulated works with a view to avoiding the FAILURE.
  The last significance of these patterns will be evaluated by an expert in the field, a marketing manager or a network administrator.
- Data Mining tools may also computerize the process of predictive information retrieval in great data bases. Questions which normally implied large analyses, may promptly find their answer in the data. A typical example of predictive problem consists in discovering the level of the targeted market sector.

Data Mining resorts to the data in the most recent editorial offers with a view to identifying the works susceptible of being consulted at a maximum level. Other predictive issues include the prevention of the failure of other forms of failure.

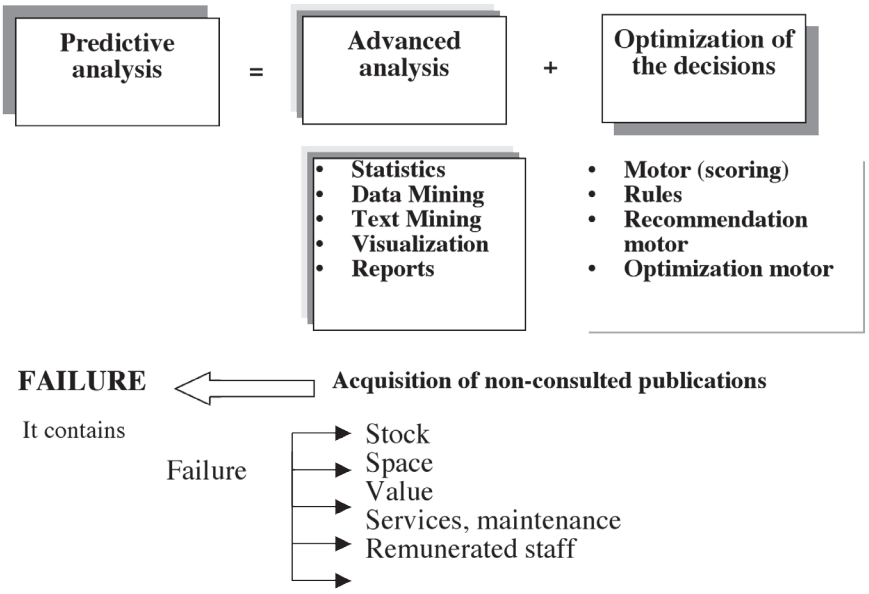


Figure 10. Elements of predictive analysis

A "data mining" application stands for an implementation of "data mining" technology, which solves a specific task or a research issue. Example of the application areas:

- A library may analyze the management of collections from the recent period in order to improve the objectives of meeting the users' needs and in order to determine what publications are accessed (have had the greatest impact) during the last months. The data will include the most often quoted works and also information with respect to the works that have been quoted and do not exist within the library. The results will be distributed to the Evidence-Acquisition department, a fact that will determine the representatives of the department to revise the recommendations from the perspective of the key attributes in the process of decision. The continual analyses of the deposited data allow for the best procedures to be applied in the specific acquisition situations.

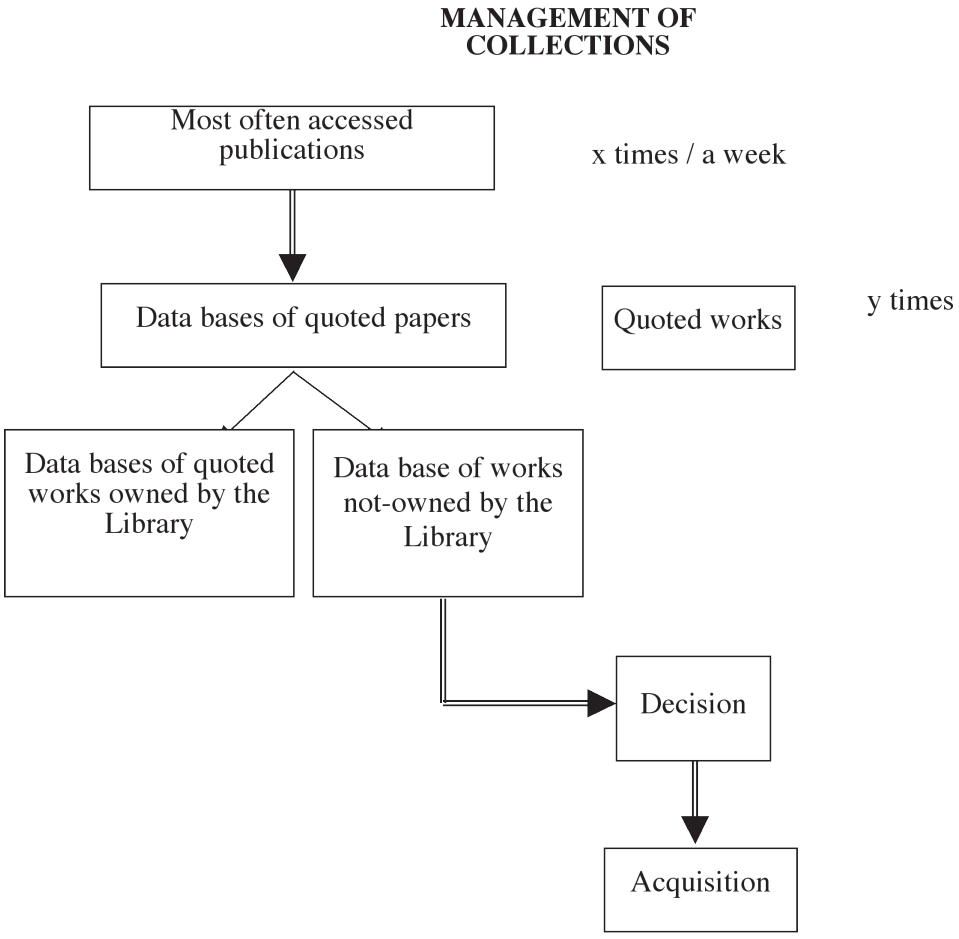


Fig. 11 Management of collections

Management of collection development

- The library may control the research activity in the respective university in order to identify the fields which should manifest the greatest interest in acquiring new publications. By resorting to data mining, there may be easily identified the most solicited authors in different fields, the professorial staff's articles appeared within our country and abroad, there may be created a data base with quotation indices.

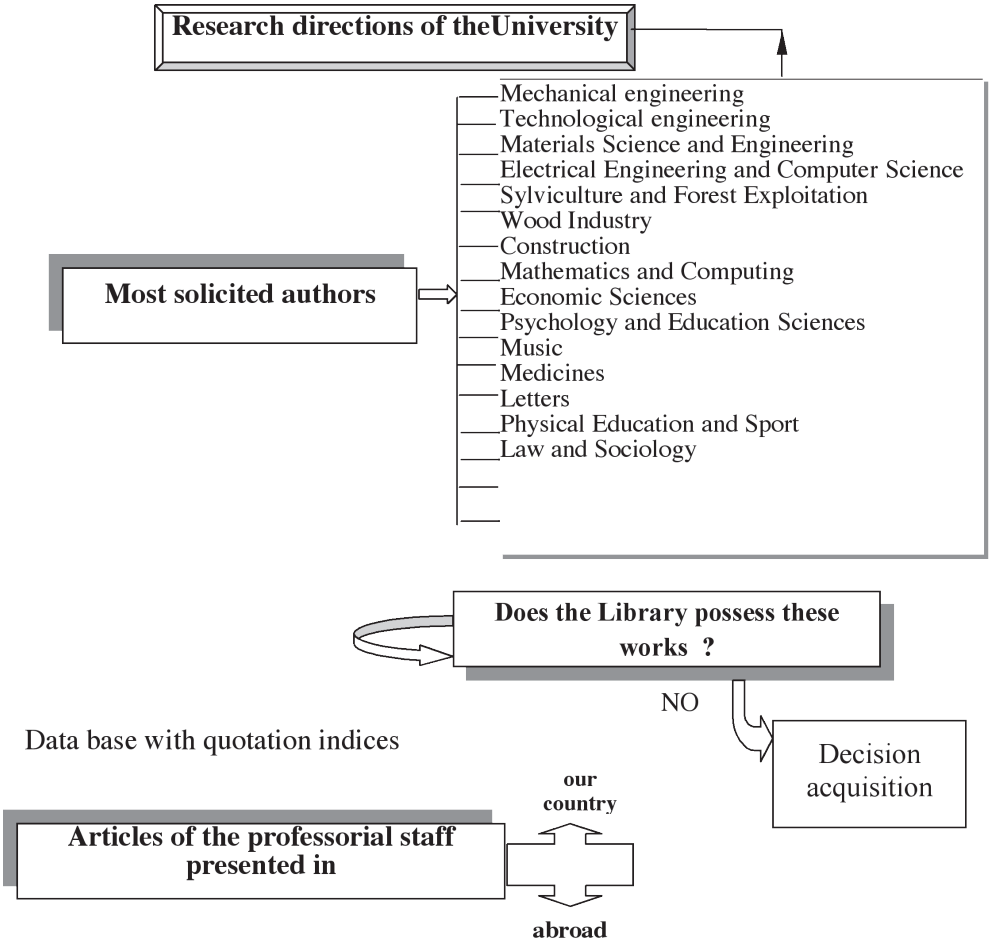


Figure 12: **Management of collections. Development and directions of research**

Data mining techniques may give up the benefits of the existing software and hardware platforms automation in order to enhance the value of the existing informational resources, and they may be implemented upon new products and systems once they are available on-line. When these techniques are implemented upon customer/server processing systems or parallels of high performance, they may analyze massive data bases and may answer questions such as: What department, field proves itself most satisfied with the offer of the library?

Data Mining technology is prepared for applications, as it is supported by three technologies that are currently sufficiently developed, which are:

1. Massive data collections
2. Powerful multi-processor computers
3. Data search algorithms

The main components of data mining technology have been under development for decades, in research fields such as statistics, artificial intelligence and learning of intelligent machines. Currently, the development of these technologies in compliance with relational data bases browsers of high performance and the ample efforts of data integration make these ones extremely practical for the fields of data depositing.

The key for understanding the different aspects of data mining consists in the distinction among the applications, operations, techniques and algorithms of data mining.

Applications	Data bases market Classification of the users
Operations	Classification and prediction Clustering Analyses of associations Prevision
Techniques	Neuronal networks Decision trees Close neighbour algorithms Naïve Bayesian Cluster analysis

Table 3: Data Mining. Elements

Operations

An application which resorts to data mining technology will implement one or more data mining operations (many times called tasks). Every operation reflects a different modality for the distinction of the patterns or of the orientations within a complex data set.

Classification and Prediction

Classification stands for the operation which is most used by the data mining commercial instruments. It is an operation that helps organizations discover

certain patterns in vast and complex data banks with a view to solve specifically business issues.

An example of classification: University Library wants to identify those users who might be interested in a new opportunity of data bases acquisition. It also launched formerly on the market such a product and it possesses data that indicate which of its users answered the previous offer. The goal consists in understanding what factors identify the users who positively answered the offer, so that the effort made in marketing and in acquisition should be focused more efficiently.

There is, in the users' registration, a field that it set as being "true" or "false" according to the users' answer to the offer. The field is denominated "target field" or "dependent variable" for classification. The goal is to analyze in what way other attributes of the users (such as the field they study, the year, the school performances, the number of the visits to the library) influence the class they belong to (the class indicated by the target field). This piece of information will be deposited in other fields from the user's registration. The distinct fields included within the analysis are identified as fields or independent or predictive variables.

Classification techniques: The manner in which "data mining" instruments analyze the data, and the type of information they offer, depends upon the techniques resorted to. The most common techniques of classification are: the decision trees and the neural networks. If a decision tree is made use of, we shall have at our disposal a set of ramified conditions which successfully divide users into groups, defined by the values from the independent variables. The aim is being able to produce a set of rules or a certain type of model.

By contrast, a neuronal network identifies the class a certain user belongs to, not being able however to state why. The factors which determine the classification are not available for analysis, remaining however involved within the network. Another set of techniques resorted to for classification are the k-nearest neighbor algorithms.

Softwares for Data Mining classification

1. Multiple approaches, typically including both a decision-tree and a neural network models, as well as some way to combine and compare them.
2. Classification with Neural networks
3. Classification with Decision tree or Rule-based approaches
4. Bayesian and Dependency Networks

5. SVM (Support Vector Machines)
6. Other approaches, including Rough sets and Genetic algorithms.
7. Analysis of classification results, ROC curves, and more

1. Software for Classification using multiple approaches

commercial: | free

- Affinium Model Suite, includes linear regression, logistic regression, CHAID, neural networks, and genetic algorithms.
- Clementine from SPSS, leading visual rapid modeling environment for data mining. Now includes Clementine Server.
- KINOsuite PR, extracts rules from trained neural networks.
- Knowledge Studio, featuring multiple data mining models in a visual, easy-to-use interface.
- MarketMiner automatically selects the best mining technique using: Statistical Networks, Logistic and linear regression, K nearest neighbors, and Decision trees (C4.5).
- MLF: machine learning framework for Mathematica, the multi-method system for creating understandable computational models from data.
- Oracle 9i Data Mining, embeds data-mining functionality into the Oracle9i database, for making classifications, predictions, and associations.
- Polyanalyst, features multiple classification algorithms: Decision Trees, Fuzzy Logic, and Memory Based reasoning.
- Predictive Dynamix Data Mining Suite integrates statistical, graphical, and ROC analysis with neural network, clustering, and fuzzy models.
- PredictionWorks, includes decision tree (gini, entropy, C4.5), logistic regression, k nearest-neighbor, naive bayes and linear regression. Free test over the web!
- Previa Classpad, provides an interactive environment for classification using neural networks, decision trees, and bayesian networks.
- prudsys DISCOVERER: non-linear decision trees (NDTs) and sparse grid methods for classification
- Purple Insight MineSet, offering several classification methods.
- STATISTICA Data Miner

free:

- JAM, Java Agents for Meta Learning (applications to fraud and intrusion detection).
- MLC++, a library of C++ classes for supervised machine learning, including multiple classification algorithms.
- SIPINA-W, produces decision graphs and trees. Includes several classification methods. (Win). Shareware

2. Neural Network Software for Classification

sites:

- Neural Network FAQ list, includes free and commercial software, maintained by Warren Sarle of SAS.
- Portal for Forecasting with neural networks, including software, data, and more.

commercial | *free*

- Alyuda NeuroIntelligence, supports all stages of neural network design and application.
- BioComp iModel(tm), self-optimizing, non-linear predictive model development.
- COGNOS 4Thought, predictive modeling tool offering Effectiveness measurement, What-if analysis, and Forecasting
- BrainMaker, a fast neural network system, now with MMX acceleration (Win 95/NT)
- KINOsuite
- MATLAB Neural Net Toolbox, a comprehensive environment for neural network research, design, and simulation within MATLAB
- NeuroSolutions, powerful and flexible neural network modeling software.
- NeuroXL, neural networks software for classification and prediction of simple and complex data in Excel.
- NeuralWorks Predict Predict 3.0 and Professional II/PLUS.
- SPSS Neural Connection 2, with Bayesian Network, Data Output Tool, model weights and more.
- STATISTICA Neural networks, comprehensive and user-friendly nn application with many charting options, network architectures and training algorithms.

free and shareware:

- NuClass7, freeware for fast development, validation, and application of neural and conventional classifiers including multilayer perceptron, functional link net, piecewise linear net, nearest neighbor classifier, self organizing map.
- Sciengy RPF(tm) for Data Mining, a free experimental Windows application self-organizing neural networks, a convenient user interface and ability to work with text data files.
- Tiberius, MLP Neural Network for classification and regression problems.

3. Decision Tree Software for Classification

commercial | free

- AC2, provides graphical tools for data preparation and builing decision trees.
- Alice d'Isoft 6.0, a streamlined version of ISoft's decision-tree-based AC2 data-mining product, is designed for mainstream business users.
- C5.0/See5, constructs classifiers in the form of decision trees and rulesets. Includes latest innovations such as boosting.
- CART 5.0 decision-tree software, multiple winners of KDD Cup. Advanced facilities for data mining, data pre-processing and predictive modeling including bagging and arcing.
- Compumine Rule Discovery System, has a complete coverage of rule-based predictive modeling methods for both classification and regression, with a user-friendly interface.
- DrawBTree, Binary Decision Tree flowcharting software, as well as examples of how these flowcharts are used in documents.
- DTREG, generates classification and regression decision trees; finds optimal tree size; supports variable costs, priors and variable weights. Download free demo version.
- Decisionhouse, provides data extraction, management, pre-processing and visualization, plus customer profiling, segmentation and geographical display.
- Fair, Isaac Model Builder for Decision Trees, advanced tree-building software that leverages your data and your business expertise to guide you in strategy development.
- KnowledgeSEEKER, high performance interactive decision tree analytical tool.
- Neusciences aXi.DecisionTree, ActiveX Control for building a decision tree. Handles discrete and continuous problems and can extract rules from the tree.
- PolyAnalyst, includes an information Gain decision tree among its 11 algorithms.
- Purple Insight MineSet, offering decision trees and other classification methods.
- Shih Tree Builder, Classifier/regression/probability tree, manual/automatic split, pruning, priors, misclassification costs, train/test parallel monitoring, mixed algorithms, exhausitive search of possible splits.
- SPSS AnswerTree, easy to use package with CHAID and other decision tree algorithms. Includes decision tree export in XML format.
- XpertRule Miner (Attar Software), provides graphical decision trees with the ability to embed as ActiveX components.

free and shareware:

- C4.5, the "classic" decision-tree tool, developed by J. R. Quinlan, (restricted distribution).
- BTreePseudo, pseudo-code for a decision-tree program.
- Classification Tree in Excel, from Angshuman Saha
- DM-II system, includes CBA for classification based on associations, and many more features.
- GAtree, genetic induction and visualization of decision trees (free and commercial versions available).
- IND, provides Gini and C4.5 style decision trees and more. Publicly available from NASA but with export restrictions.
- LMDT, builds Linear Machine Decision Trees (based on Brodley and Utgoff papers).
- Mangrove, a tool for visualization of decision tree learning,
- OC1, decision tree system continuous feature values; builds decision trees with linear combinations of attributes at each internal node; these trees then partition the space of examples with both oblique and axis-parallel hyperplanes.
- ODBCMINE, analyzes ODBC data bases using C4.5, and outputs simple IF.. ELSE decision rules in ascii.
- PC4.5, a parallel version of C4.5 built with Persistent Linda (PLinda) system.
- SMILES, advanced decision tree learner, with new splitting criteria, non-greedy search, extraction of different solutions, boosting, cost-sensitive learning, and more.
- Random forests from Leo Breiman, a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest.

Rule-based approach

commercial:

- Datamite, enables rules and knowledge to be discovered in ODBC-compliant relational data bases.
- PolyAnalyst, supports decision tree, fuzzy logic rules, and other classification algorithms.
- SuperQuery, business Intelligence tool; works with Microsoft Access and Excel and many other data bases.
- WizWhy, automatically finds all the if-then rules in the data and uses them to summarize the data, identify exceptions, and generate predictions for new cases.
- XpertRule Miner (Attar Software) provides association rule discovery from any ODBC data source.

free:

- CBA, mines association rules and builds accurate classifiers using a subset of association rules.
- CN2, inductively learns a set of propositional if...then... rules from a set of training examples by performing a general-to-specific beam search through rule-space.
- KINOsuite-PR extracts rules from trained neural networks.
- PNC2 Rule Induction System, Windows software tool that induces rules using the PNC2 cluster algorithm.

4. Bayesian and Dependency Networks Software

Overview pages | commercial | free

- Kevin Murphy's Bayesian Network Software Packages page
- Google's list of Bayes net software.

commercial:

- AgenaRisk, visual tool, combining Bayesian networks and statistical simulation (Free one month evaluation).
- Analytica, influence diagram-based, visual environment for creating and analyzing probabilistic models (Win/Mac).
- AT-Sigma Data Chopper, for analysis of data bases and finding causal relationships.
- BayesiaLab, complete set of Bayesian network tools, including supervised and unsupervised learning, and analysis toolbox.
- Bayesware Discoverer 1.0, an automated modeling tool able to extract a Bayesian network from data by searching for the most probable model
- BNet, includes BNet.Builder for rapidly creating Belief Networks, entering information, and getting results and BNet.EngineKit for incorporating Belief Network Technology in your applications.
- DXpress, Windows based tool for building and compiling Bayes Networks.
- Ergo(tm), Bayesian Network Editor and Solver (Win and Mac demos available)
- Flint, combines bayesian networks, certainty factors and fuzzy logic within a logic programming rules-based environment.
- HUGIN, full suite of Bayesian Network reasoning tools
- KnowledgeMiner , uses self-organizing neural networks to discover problem structure (Mac platform)
- Netica, bayesian network tools (Win 95/NT), demo available.
- PrecisionTree, an add-in for Microsoft Excel for building decision trees and influence diagrams directly in the spreadsheet

free:

- BAYDA 1.0
- Bayesian belief network software (Win95/98/NT/2000), from J. Cheng, including
  BN PowerConstructor: An efficient system for learning BN structures and parameters from data. Constantly updated since 1997.
  BN PowerPredictor: A data mining program for data modeling/classification/prediction. It extends BN PowerConstructor to BN based classifier learning.
- Bayesian Logistic Regression Software, for large-scale Bayesian logistic regression (Windows and Linux)
- Bayesian Network tools in Java (BNJ): an open-source suite of Java tools for probabilistic learning and reasoning (Kansas State University KDD Lab)
- FDEP, induces functional dependencies from a given input relation. (GNU C).
- GeNIe, decision modeling environment implementing influence diagrams and Bayesian networks (Windows). Has over 2000 users.
- JavaBayes
- jBNC, a Java toolkit for training, testing, and applying Bayesian Network Classifiers.
- MSBN: Microsoft Belief Network Tools, tools for creation, assessment and evaluation of Bayesian belief networks. Free for non-commercial research users.
- PNL, Intel Open-Source Probabilistic Network Library
- Pulcinella, tool for Propagating Uncertainty through Local Computations based on the Shenoy and Shafer framework. (Common Lisp)

5. SVM (Support Vector Machines) Software for Classification

commercial:

- EQUBITS Foresight(tm), SVM-based Predictive Modeling.
- KXEN, Components, based on Vapnik's work on SVM.

free:

- BSVM, a decomposition method for support vector machines (SVM) for large classification problems.
- Gist, web interface to SVM
- Kernel Machines and related methods website
- LS-SVMlab, Least Squares - Support Vector Machines Matlab/C Toolbox
- LIBSVM, a support vector machines (SVM) library for classification
- pcSVM, a framework for support vector machines, including a pcSVM demo program which trains SVM on generated 2-D classification problems and draws the computed decision line together with the identified support vectors.

- SVM-light, popular implementation of Support Vector Machines (SVMs) by Thorsten Joachims.
- SvmFu 3, SVM package in C++, under GPL (source only, requires compilation)

6. Genetic Programming, Rough Sets and other classification software

commercial:

- Datalogic, professional tool for knowledge acquisition, classification, predictive modelling based on rough sets.
- Evolver, genetic programming
- GAtree, genetic induction and visualization of decision trees (free and commercial versions available).
- Genalytics GA3, uses genetic programming to dynamically build predictive models.
- gepsoft APS 3.0, an extremely flexible modeling tool for Function Finding, Classification, and Time Series Prediction based on Gene Expression Programming; generates Visual Basic; C++ and more.
- MARS, J. Friedman's automated logistic regression for binary classification problems. Automatic missing value handling, interaction detection, variable transformation.
- WINROSA, automatically generates fuzzy rules, based on the fuzzy ROSA method (Rule Oriented Statistical Analysis).

free:

- Grobian, user-friendly software to analyse data with rough set technology (C++).
- Rough Set Exploration System (RSES), contains classification based on rough sets, decision tree, LTF-network, instance based classification and data discretization (free for non-commercial use).
- TiMBL 2.0, nearest neighbour approach

7. Software for Analysis of Classification

- Analyse-it, Method evaluation & validation software for Excel. Includes ROC curves, method comparison, NCCLS imprecision, reference ranges, and NCCLS linearity. Free 30-day trial.
- MedCalc, Windows statistical program for biomedical researchers; includes ROC analysis and more.
- ROCOn, a tool to aid ROC analysis of machine learning classifiers.

Comprehension and Prediction: Sophisticated classification techniques support us in retrieving new patterns in large and complex sets of data. Therefore, classification stands for a powerful help in comprehending a certain problem, even if it represents the installment of the answers to a promotion of the library services.

Under certain circumstances, it provides an improved comprehension. This may suggest new initiatives and may provide information which improves the process of decision-making in the future. However, in many cases, the reason for the development of a precise classification model consists in improving the prediction ability.

They say a classification model should be accomplished in compliance with historic data, in which case the result for every registration is known. This is subsequently applied to a new set of non-registered data, with a view to predict the result for every registration.

There are significant differences between classifying data with a view to understanding the existing users' behavior and resorting to the respective classification in order to predict the future behavior.

As for sets of data from the past, it is often possible to produce a set of rules or a mathematical function which should accurately describe every registration.

Building a good predictive model implies avoiding the over-charge by testing and harmonizing the model, in order to ensure the fact it may be generalized to the new data.

Clustering

Clustering stands for an operation that cannot be controlled. It is resorted to where we want to retrieve groups of similar records in the data we have at our disposal, with no other precondition the respective similarity implies. Clustering is used in order to identify the groups of interest from a data base of the users, groups which have not been previously recognized. For example, it may be used in order to identify similarities in the users' information needs, with a view to inventing and selling new services of documentary information.

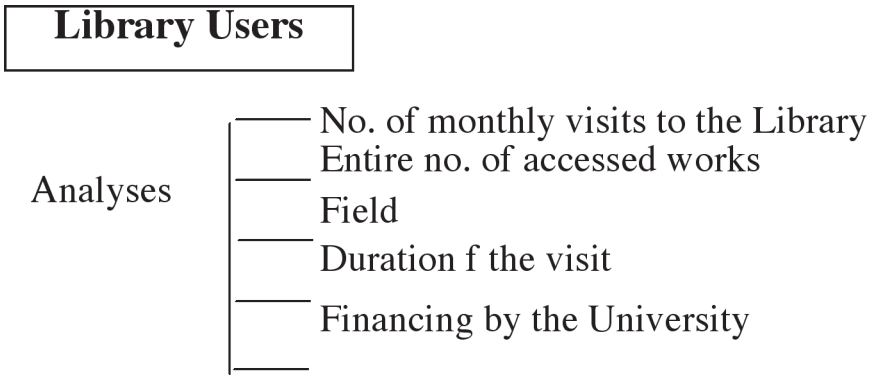
Clustering is obtained, as a rule, by means of statistic methods, such as the k-means algorithm, or a special form of the neuronal network called "Kohonen network map". Whatever should be the method resorted to, the basic operation remains the same. Every registration is compared to a set of existing clusters, which are defined by their "center". A certain registration is due to the cluster that it is nearer to, and the former changes the value which defines the cluster. Several steps are made until a set of data has re-allotted themselves the records and has modified the centers of their clusters until the discovery of the optimal solution.

For example, searching for clusters among the library users, there might be needed to analyze more factors, among whom the number of monthly visits to the library, the whole number of the accessed works during one visit, the time of the visit and the financing by the university.

Clustering is often interpreted as an exploration exercise before continuing to search the data by means of the classification technique. For this reason, a good visual knowledge represents a further support for clustering: it makes you deepen into the work with the clusters, with a view to recognizing if the identified clusters make sense and if they can be of help in the context of the business.

Classification technique

Clustering



Analysis of the Association and Sequential Analysis

The analysis of the association stands for an uncontrolled form of "data mining" which searches for connections among the registrations from a set of data. The analysis of the association is sometimes defined as "analysis of the accessed sites, which is its most used application. The aim is to discover, for example, what field is mostly accessed.

Prognosis

Prognosis is often undertaken with regression functions – statistic methods for the examination of the relations among the variables in order to predict the next value. The statistic packages, such as SAS and SPSS, provide an extensive range of such functions which may manipulate more and more complex issues.

Nonetheless, such statistic functions usually imply a significant knowledge of the techniques resorted to and of the original conditions applied to their implementation. Data mining tools may also provide functions for prognosis. In particular, the neuronal network was used upon vast areas for the prognosis of the stocks on the market.

An important distinction may be operated between two different types of the prognosis issue.

The simpler problem consists in the prognosis of a single continuous value on the basis of series of unordered examples.

Techniques

An operation of the type data mining may be obtained by resorting to a number of techniques or methods.

Every technique may be itself implemented in different manners, by resorting to different algorithms.

Clustering Algorithms

The analysis of the clusters stands for a process of identification of the clusters existing between articles on the basis of the resemblance and distinctions among them. Unlike classification, clustering does not imply the previous identification of a target variable. An algorithm checks the potential groupings from the multitude of data and endeavors to obtain an optimal delimitation of the articles on the basis of those groupings.

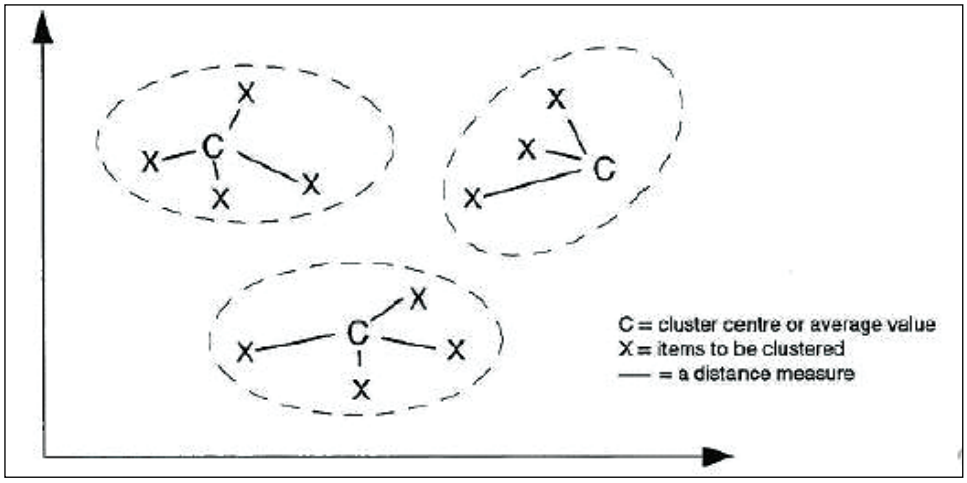


Figure 14: Grouping of the Clusters

Clusters are usually placed around a "center" or average value. The original manner of defining and regulating the centers varies according to the algorithms. One manner consists in beginning with a random set of centers, which are subsequently regulated, added and eliminated while the analysis advances.

In order to identify the articles from a cluster, we must resort to a sort of measure which equates the resemblance among the articles in a cluster and the distinctions from the articles in other clusters. The resemblance and the distinctions among the articles are usually measured as the distance between them and the others and to the centers of the clusters from the multidimensional space, where every dimension stands for one of the variables which undergo the comparison.

Nearest Neighbour

The nearest neighbor (more exact k-the nearest neighbor, as well as k-NN) stands for a prediction technique adequate for the classification models.

Unlike other predictive algorithms, the stimulation data are not scanned or processed for the creation of a model. In exchange, the stimulation data stand for the model. When a new case is presented, the algorithm operates a search in all the data, in order to find a subset of cases which mostly resemble one another, and uses it in order to predict the consequence.

There are two main drivers in the algorithm k-NN: the number of the nearest cases which are to be used (k) and a metrics in order to measure what really means the nearest.

Every use of the algorithm k-NN implies what we call entire positive value of k. This determines how many cases already existent have been already studied when a new case is predicted. K-NN refers to a family of algorithms that can be understood as 1-NN, 2-NN, 3-NN and so on and so forth. For example 4-NN indicates that the algorithm will make use of the nearest 4 cases in order to predict a new case.

k-NN is based on the concept of distance and this implies a metrics in order to determine the distances. All metrics have to result in a specific number as for the comparisons. Any metrics resorted to is arbitrary and extremely important. It is arbitrary because it exists no control definition of what is called a "good" metrics. It is important because the choice of a metrics has a strong impact upon the predictions. This means an expert in business is needed in order to contribute to the determination of a good metrics. In order to classify a new case, the algorithm calculates the distance from the new case to each case (line) of the stimulation data. It is predicted for the new case to have the same consequence as the prevailing consequence in the k close cases from the stimulation data.

Neuronal Networks

A key difference between the neuronal networks and any other technique is that neuronal networks operate exclusively on the numbers. It results that any non-numeric data in independent columns must be converted into numbers before being able to use the data in a neuronal network.

Naïve-Bayes

Naïve-Bayes constitutes a classification technique that is both predictive and descriptive. It analyzes the relation between every independent variable and every dependent variable in order to derive a conditional probability for every relation.

Naïve-Bayes implies a single examination of the learning set in order to generate a model of classification, which makes it the most efficient data mining technique. However, Naïve-Bayes does not operate with continuous data, so as any variable, independent or dependent, which possesses continuous values, has to be incapsulated.

Decision Trees

Decision trees constitute one of the most used data mining techniques. They are easy to use, the results are intelligible by a regular user, they address a wide range of classification issues and they manifest efficacy in processing o large volume of data.

An algorithm based on decision trees divides the set of data with a view to building a model that classifies every registration in terms of target field or variable. An example consists in a decision tree which divides a set of data according tot the way the users did or did not access a certain site.

The most used algorithms on the basis of decision trees are CHAID, CART and C4.5. CHAID (Chi-square automatic interaction detection) and CART (Classification and Regression Trees) were developed by statistics. CHAID may produce an tree with multiple sub-knots for every division. CART implies less preparing of the data than CHAID, but it always divides the set in only two parts. C4.5 comes from the universe of the intelligent machines capable of learning and it is based upon the theory of information.

In order to generate a decision tree out of a set of stimulation data, it is necessary to progressively divide the data in smaller sub-sets. Every iteration takes into consideration the data from a single knot. The first iteration processes the root-knot, which contains all the data. The iterations which follow work upon derived knots which will contain sub-sets of data. In every iteration we must choose that independent variable which divides the data with most efficacy.

This means the produced sub-sets must be as "homogenous" as possible towards the dependent variable.

V Text Mining Applied in Digital Library

The information in electronic format has become omnipresent. The Internet provides its users with large amounts of information in any field of interest. Within this context the apparition and development of systems assisting and facilitating the access to information has become compulsory.

These systems may be of real help with regard to the automated classification of documents, information, information extraction, summarizing, browsing efficiency, increase of machines understanding, etc.

The greatest part of these systems is framed in the field of Text Data Mining. The information extraction is one of them, having a great importance in the nowadays context.

Web application

Through this project we aimed at conceiving a programme of knowledge extraction in HTML documents. The program recognizes events of a certain type (weather, sport, politics, text data mining, etc) taking into account the way it will be driven (the concept-based dictionary it chooses). These events may be provided to the user or the whole context where the event occurred can be extracted in order to indicate the initial form where the event was framed.

This project aims to be of great help in case the information is searched in more web pages (Yahoo, Google, CNN site, etc.) for a certain area and a certain period when the project can be useful. For each type of events (conferences, specialized education, higher education) there is a concept-based dictionary. Conceiving such a dictionary is a very demanding task. We are endeavoring to think up an extensible system to allow the extraction of specific events. At present the automated built dictionaries are of greatest interest, through learning algorithms.

Definitions and acronyms

Text Data Mining = TDM
Concept-Based Dictionary = DC
Part of Speech Tagger = POST
Text Data Mining = Field of extracting new items of knowledge from the text
Concept = assembly of all items of information of a certain type that follows a certain syntax imposed by concept (ex: {**information science**, university, sem. I}, {**information science**, university RO, sem.II}, {**information science**, X, Y, for any X and Y valid, plus a series of valid syntaxes valid for the concept of information science and attributes X and Y)
Concept-based dictionary = Assembly of concepts for a particular field
POST = Applications having as aim the annotation of a text with specific speech parts specific to each word in the sentence
WordNet = A data base with English language words, with classification at semantic level
Event = an instance of a concept, that is an assembly of information that verifies – respects the structure imposed by a concept (ex: {**information science**, university, sem.I})
Segment = part of a text that represents a sentence or a phrase

General outlines

The field of events extraction from text with natural language (for example Web pages) is in continuous development. The researchers' meetings in the field were of great interest, one of the most important one being MUCs (Message Understanding Conference). The systems developed and tested within these conferences were oriented towards the discovery of terrorist messages, of attacks upon messages on the Internet. Taking into account the fact that the text is in natural language, which has an increased complexity, and, therefore, a formalization is almost impossible, the results obtained in the correct identification of events is quite good, raising up to 80-90% in some cases.

Description upon component parts
Diagram moduli

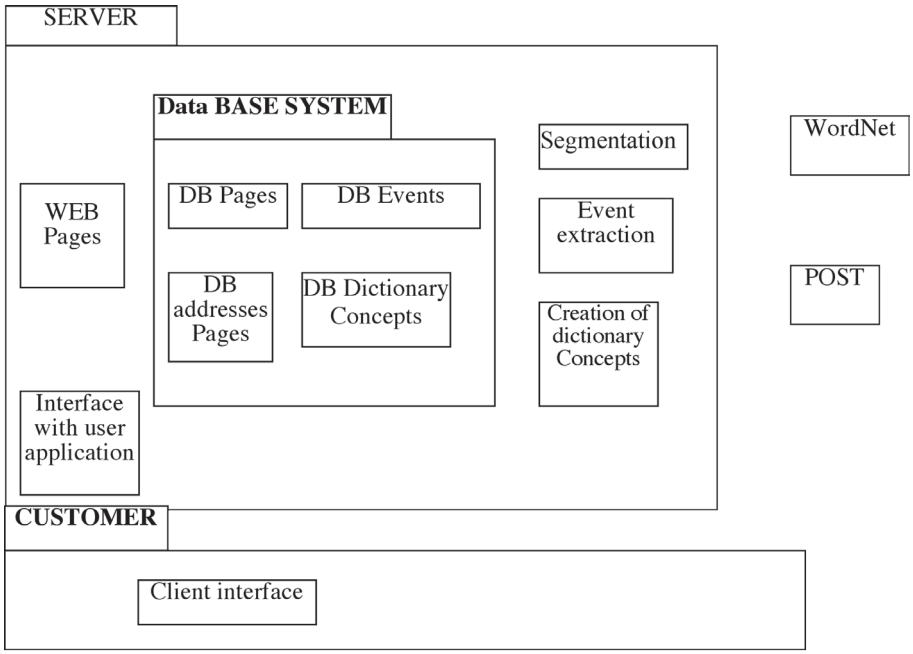


Figure 15: Components moduli diagram

Modules and their components

MODULE	FUNCTION
Pages Addresses Data Base	It keeps the list of Web pages for a certain field on which the events are searched
Pages Data Base	It keeps the Web pages, indexed and on fields categories
Events Data Base	It keeps the identified events, indexed according to fields, time locations and other criteria
Concept-based Data Base	It keeps the concepts that will be searched and indexed on fields
Web pages fetching	It fetches Web pages from Pages Addresses Data Base and it stores them in the Pages Data Base. This module may recursively fetch the pages indicated by links in the current page, from the same server, till a certain fetching depths
Segmentation	It divides the text from the Web pages within the Pages Data Base in segments
POST	This module receives the segments and annotates them with the speech parts corresponding to words in segments
Events extraction	It extracts events from annotated segments, received from the POST module, on the basis of concepts in the Concept-based Dictionary and of WordNet, adding them to the Events Data Base
Concept-based Dictionary	It creates the concepts specific to a certain field, through learning algorithms, then it adds them to the Concept-based Data Base
WordNet	It provides relationships between words of semantic nature
User interface	It receives requests for events searching, requests received from the Events Data Base
Customer interface	It assures the user interface, the introduction of events queries, options setting, results display according to various criteria

Table 4: Component moduli

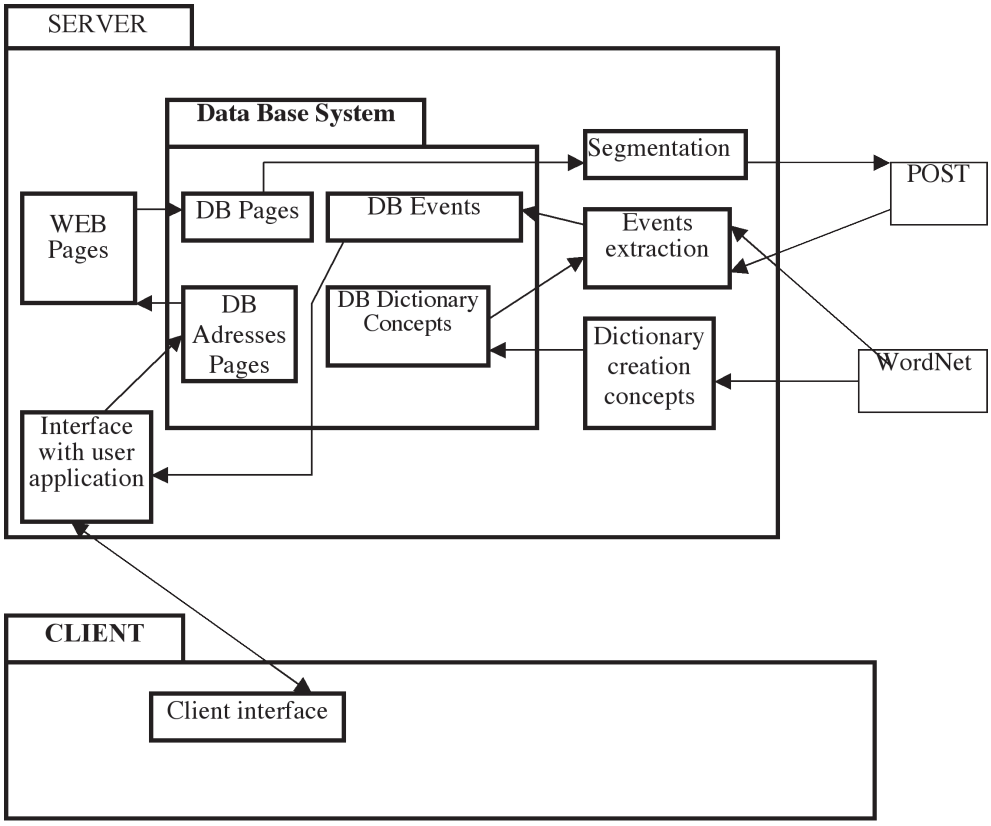


Figure 16: Moduli diagram

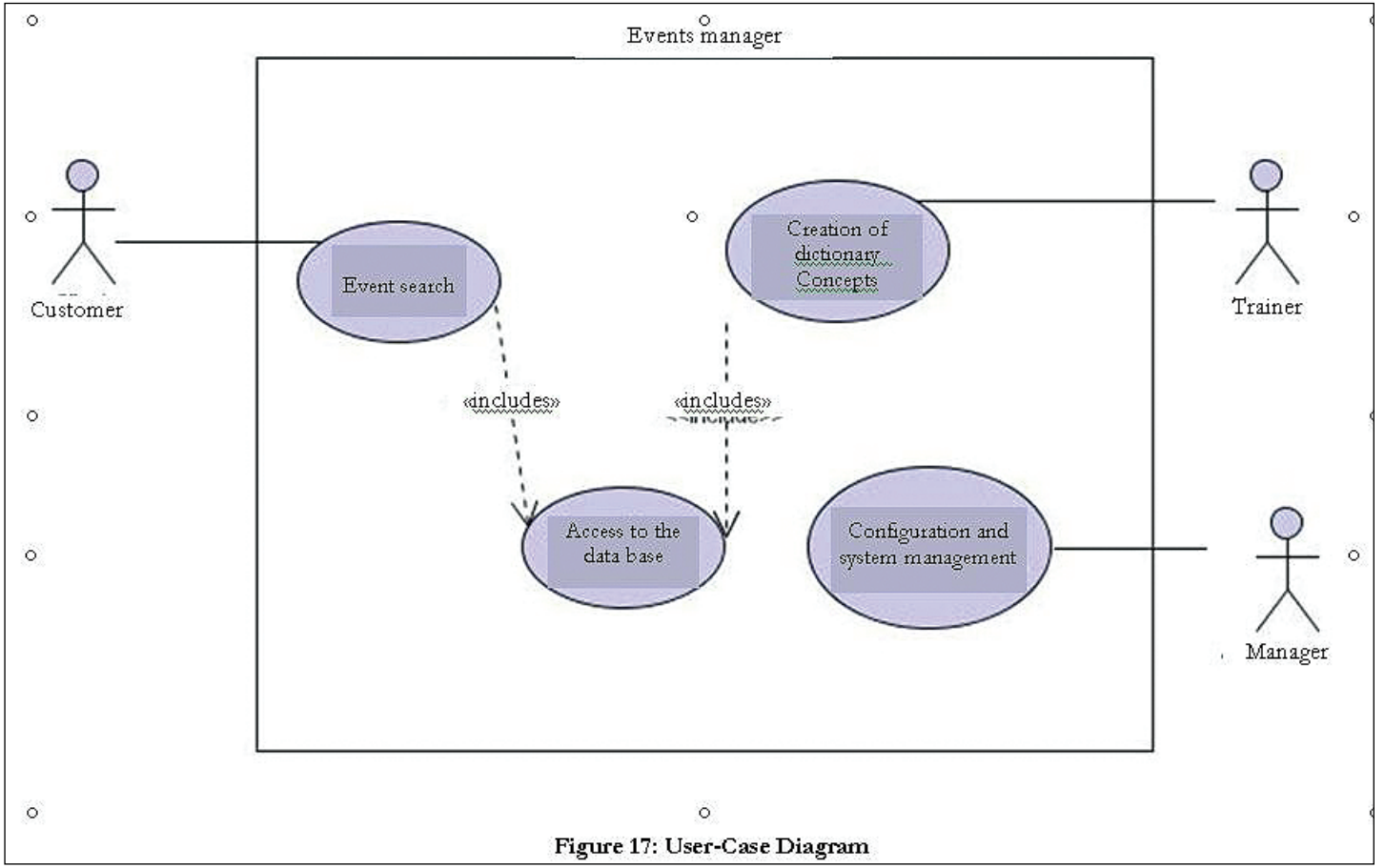


Figure 17: User-Case Diagram

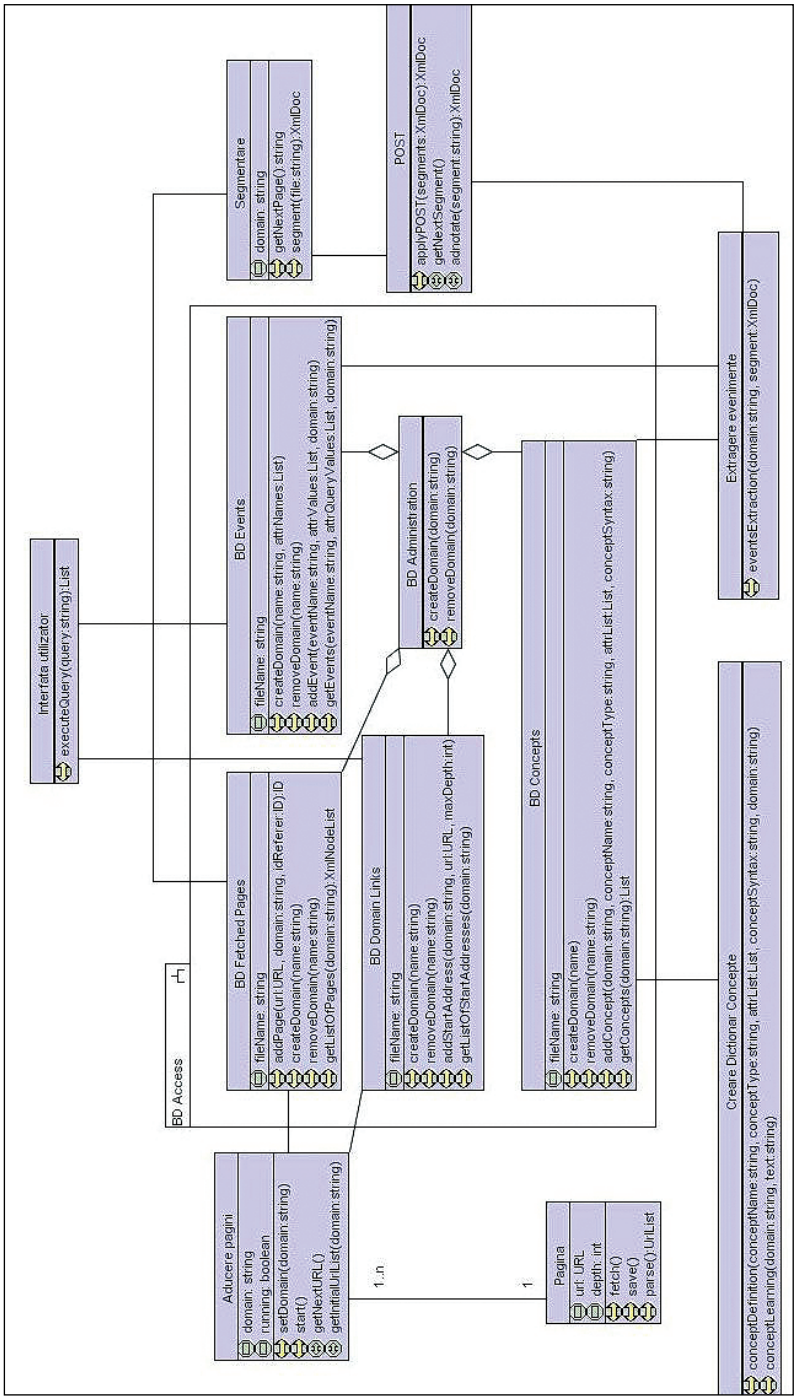


Figure 18: Moduli interface description – classes diagram

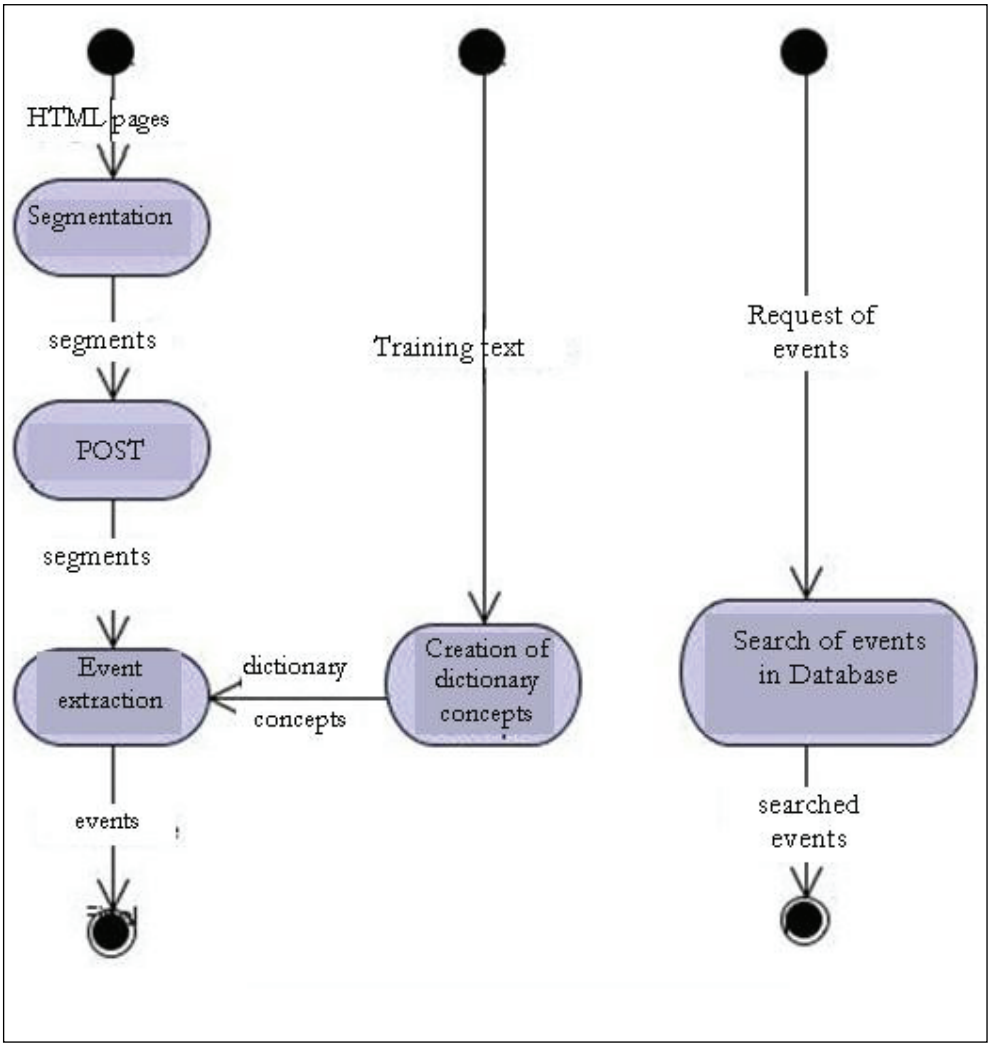


Figure 19: Detailed Description Design

Detailed description of modules

Pages Addresses Data Base:

– it corresponds to “BD Domain Links” class
– it is a XML Data Base
– it keeps the list of start pages addresses
– only the start pages are kept in this database, but the fetching module searches and fetches the pages making reference to start pages found on the same server, for a certain fetching depth

```
<xml>
    <domain name=”_tiin_a inform_rii”>
        <urlname=”www.yahoo.com/information science”  adancime=3>
        <url name=”www.cnn.com/ information science ”  adancime=2>
                                    …..
    </domain>
</xml>
```

Pages Data Base:

– it corresponds to BD Fetched Pages class
– it is a XML Data Base
– it keeps the index of fetched pages in XML format, and the pages in distinct files
– only the pages containing text are fetched and kept, not those indicated by links to images or to other insignificant types for application.
– For each page there is an index in the Pages Data Base that contains information about the page url, the last updating date, the name of the file where it was stored, a unique id of each page and the list of pages ids with which a page is in relation.

Ex :

```
<xml>
    <domain name=”_tiin_a inform_rii”>
        <pagina url=”www.yahoo.com/information science /” lastUpdated=”22.11.2002”
            fileName=”information science_com_yahoo_www” id=1>
    </pagina>
    <pagina url=”www.yahoo.com/_tiin_a inform_rii/europe”
          lastUpdated=”22.11.2002”
        fileName=”europe__tiin_a inform_rii_com_yahoo_www” id=2>
    <upPage>1</upPage>
    </pagina>
  </domain>
<xml>
```

Events Data base:

– it corresponds to BD Events class
– it is a XML Data Base
– it keeps the identified events, indexed according fields, time locations and other criteria
– Concept = assembly of all items of information (events) of a certain type that follows a certain syntax imposed by concept (ex: {Bucuresti, curriculum}, {information science, Brasov, University, {information science, X, Y}, for any X and Y valid, plus a series of syntaxes valid for the concept of information science and attributes X and Y)
– The following items of information must be kept for the concept:
– Name of concept
– Type of concept
– List of attributes
– Syntax of concept – positioning the attributes related to a certain concept

Web pages fetching:

– It corresponds to "Pages fetching" class
– It fetches Web pages from Pages Addresses Data Base and it stores them in the Pages Data Base. This module may recursively fetch the pages indicated by links in the current page, from the same server, till a certain fetching depth.
– This module appeals to an external application, specialized on web pages recursive fetching; yet, it can fulfill this function internally.
– It contains analysis function of the page to extract the new links, plus functions of information removal from page (ex: scripts) that are not relevant for the application

Segmentation:

– It corresponds to Segmentation class
– It divides the text from the Web pages within the Pages Data Base in segments
– A segment is a part of a text that represents a sentence or phrase and can be regarded as an entity of atomic information
– The POST applications accept this type of segments as input data
– The segmentation is made on the basis of HTML tags that helps to their delimitation (ex: <P>
 etc.) as well as text delimitation elements.

POST:

– it corresponds to “POST” class.
– this module receives the segments and annotates them with the speech parts corresponding to words in segments.
– The module interfaces the program with a Part of Speech Tagger application, which receives a segment and annotates it.

Events extraction

– it corresponds to “Events extraction” class.
– it extracts events from segments annotated, received from the POST module, on the basis of concepts in the Concept-based Dictionary and of WordNet, adding them to the Events Data Base.
– This module is one of the most complex in the system, together with the Concept-based Dictionary and Segmentation Modules.
– A concept is made up of a trigger and a series of attributes that are found as related to the trigger in a certain pattern (that implies a certain syntax).
– It uses algorithms of patterns matching to identify the possible attributes that are subsequently checked by using the WordNet-based information.

Concept-based Dictionary

– it corresponds to Concept-based Dictionary class.
– it creates the concepts specific to a certain field, through learning algorithms, then it adds them to the Concept-based Data Base.
– The Concept-based Dictionary is conceived through either manual identification of representative concepts for a certain field, these being provided to the module to be introduced in the data base, or use of algorithms of learning the concepts on certain learning pages.
– The model with learning algorithms leads to the extraction of concepts that best fit and undertake a good identification of events of similar pages to those from which learning was made.
– The Concept-based Dictionary is one of the most important components within the context of events extraction.

User interface:

– it corresponds to “User interface” class.
– it receives requests for events searching, requests received from the Events Data Base.
– the events are only received from the database, their searching and identification being concluded separately from the Event Extraction module.

VI Scientific Research and University Library
-proposition of research laboratory for information science-

University library possesses a significant quantity of data. How shall we benefit from these informational resources? Which is the role of the library?

University library constitutes an integrant part of the didactic and research process. The importance of the library within the didactic process has been exposed and dealt with in many works of specialty. In this day and age the e-learning process and e-book technology have been analyzed and optimized, the library playing a fundamental part.

Which is the role of the library in the scientific research? How does the library contribute to the visibility of the university research? How do we develop the management of the acquisition in order to support the scientific research?

Which are the projects of scientific research we can develop in the field of information science?

Within "Transilvania" University of Brasov we intend to organize a Laboratory of Informational Resources Exploitation within knowledge society.

Description of the Proposed Laboratory

The priority field that the laboratory set of themes affiliates to is that of informative and communication technology.

In the framework of the general direction, the Informational Society of the European Commission, one of the priorities aimed at consists in the increase of the functioning efficacy of the great cultural effects depositaries by means of modern techniques of management and interfacing.

The main objective of the laboratory is the creation of an evaluation and certification structure for the systems of exploitation of the informational resources within the digital library. The object is for this structure to elaborate the validation and verification of some cybermetric, bibliometric and scientometric mechanisms in order to exploit the information in the testing environment: the digital library. The structure created in guise of advanced testing laboratory will generate accredited reference materials for: advanced tests for the exploitation of the information – bibliomining – applied to the management of the digital library collection development; the systemic administration of information; holistic measurements for the library services; visibility of the impact of science in the academic community and economic environment.

The structure that will have been created will be revolutionary, datamining or data exploitation having been successfully used in medical computing, in banking, by the government organizations with a view to combating with crimes,

by ecologists and biologists with a view to discovering the pollution sources and so on and so forth.

For the first time it is created an informational structure which intends to elaborate a set of scientometric, bibliometric and cybermetric indicators, scientifically elaborated and tested, which will be used for research in information science.

The interweaving of the 5 fields – biblioteconomy, sociology, linguistics, engineering and computing – and the cooperation among the researchers will stimulate the accessibility, the information circuit for the knowledge economy and the knowledge society will provide new research possibilities if the access is based upon content and semantics, upon extraction and indexation of the information.

LERISC intends to elaborate at a national level informational models, indicators, indicators correlated to the following goals: Development of the bibliomining model. Theoretical concept. Applicative model; Information warehouse (CSI); Statistic models applied to bibliomining; Training of qualified specialists in evaluating, auditing and certifying the conformity for the informational products.

The objective consists in the researchers' regrouping, in the reunion of the competencies and in the enhancement of the resources with a view to stimulating the research in the information and communication sciences, with a view to extending their visibility and development and to the elaboration of cybermetric, bibliometric and scientometric indicators.

Within the informational society, the concepts of information and communication have acquired a new dimension: information and communication constitute an epistemological and pragmatic continuum which different specific research is placed upon.

LERISC intends to work upon the instruments and methods of access to the information, upon the consequences of the information and communication technologies with respect to the way society goes and to the new forms of sociability and to the generation of an accredited structure in the national information system which should generate informational models.

In **Romania** the reference points and the standards which will constitute the goal of LERISC are currently accomplished in a random and artisan manner, bibliometric and scientometric studies are resorted to without a previous analysis, verification and certification. It is necessary for these operations to be accomplished in an organized framework, in compliance with the European standards, a fact which will ensure the integration of the structure conceived to this purpose in the thematic area according to PC7 and to the goals of the access to the information, dissemination of the information, improvement of the

scientific research and visibility of the Romanian research.

The laboratory will stand for the first research ring chain in this field, research methodologies and approach and analysis proceedings will be elaborated with a view to creating the data warehouse for bibliomining. The development of the knowledge representation models extracted from the textual data, the automatic processing of the textual information remains a major challenge for informational society. The indicators elaborated through these techniques will be used either with a view to administrating the scientific research promotion (bibliometrics and scientometrics), or with technological and concurrential supervision (bibliometrics traditionally applied or bibliomining applied to the semantic web).

Scientific Presentation of the Laboratory

From the presentation of the existing situation it may be noticed that the goal proposed covers an area of preoccupations which IS NOT CURRENTLY COVERED by other similar preoccupations in the country.

University libraries possess integrated library systems. The Library of Transilvania University possesses the ALICE soft, which generate the data base of the library and the on-line catalogue. There has been proposed the acquisition of the LIBERTY soft, which constitutes the superior variant of the ALICE programme, which includes a full-text browser, which allows a total personalization, both introducing the data and for typing the reports and statistics, which possesses complex security rules, which may be administered and used through the intermediary of a browser, from any computer with access to the Internet, anywhere in the world, on the basis of an individual account.

By means of datamining there will be elaborated advanced tests applied to the resources: the database of the publications, the interlibrary loan, the number of the publications that have been accessed, the frequency of apparition of the works that have been accessed, the database of the users etc.

Datamining stands for the automatic practice of search for patterns in big data warehouses, for the extraction of non-trivial information, not previously known and potentially useful from the data, the science of useful information extraction from amounts of data.

The directions of research in exploiting these data are:

- Discovery of associations among objects;
- Grouping of the objects in sets of similar objects (clustering);
- Classification of the objects on the basis of their properties;
- Evaluation of the interest of the facts and properties that have been discovered;
- Preparation of the data (purification, discreetization etc)

The practical competencies necessary to the team of researchers who are involved in the accomplishment of the project:

Relational data bases, SQL and its use in C++, JAVA and other languages; Algorithms which work with a variety of data structures; Administration of the data warehouses.

Theoretical models in the field of mathematics which will be used: Clustering: metric spaces, linear algebra and functional analysis; Classification: theory of information, graphs; Latice rules, association; Theory of information; Probabilities, statistics. The notion of clustering is important for the condensation of data (concise presentation of data), identification of the tendencies within data.

In order to settle the indicators deemed to be obtained by LERISC several algorithms will be made use of:

1. Incremental clustering which is characterized by the formation of groups, successively adding objects to the groups (clusters) or forming new groups; **2.** Clustering as partitions, aiming at the determination of the groupings from their connections with the partitions induced by the attributions; **3.** AMICA (A Metric Incremental Clustering Algorithm); **4**. Decision trees.

Techniques resorted to with a view to obtaining the indicators which make the object of LERISC laboratory: Rules of induction; Neuronal networks; Conceptual clustering; Associative discovery;

After the optimization there will be elaborated the model of bibliomining and the indicators of performance proposed for the testing.

- There will be applied measures of cumulative assessment of the library information services. The measurements will be characterized by a holistic perspective, the variables resorted to being: the perspective of the measurement and the subject of the measurement.
- There will be approached the data bases of the library collection and of the users in the manner of the bibliomining archaeological model. There will be generated matrices and indices for the identification of the opportunities and for the orientation of the scientific research.
- There will also be approached the web space through the advanced retrieval system of the information, through statistic and lexical approaches of the textual data.
- There will be elaborated a model of representation of the information extracted from the textual data, which will comprise elements of semantics for the management of the scientific research prognosis (bibliomining, bibliometrics, scientometrics) and a concurrential technological supervision (bibliomining applied to the patent of information).

The generated selective pattern will be characterized by retrieval indicators advanced for testing.

In this respect there will be acquired: **the soft e-Reference** represents an interrogation archive for reference demands placed by the users of the system. The users may search full text in the "knowledge archive" created this way and they may be granted access, according to the security level of the demand, to the answer-solution offered by the reference librarian. This application may be used together with Liberty3 (file of common users) or independently; **the soft Z39.50** which allows "web cataloguing" resorting to any compatible Z39.50 catalogue in the world; the selected records are transferred in Liberty3 by a single click; **LIRIX** – system of mono and multi-linguistic information retrieval; **SPSS** may rapidly generate information of use in the decisional process and offers the possibility of presenting the results with high quality charts and of communicating the results by a variety of reports, inclusively by publication on website. All these empower us to take intelligent decisions, to rapidly discover key factors, patterns and trends in the data. SPSS is used for datamining and for the analysis of the data bases, for marketing studies and any kind of research, as SPSS is the best software in solving business and research problems resorting to statistics. SPSS is a modular line for products completely integrated for the analytical process – planning, data collection, access, data preparation and management, analysis, report drawing up and result presentation. The graphic interface makes it easy to use and it offers all the methods of data management, analysis and presentation in reports you need in order to accomplish even the most intricate analyses. **Clementine 10.**

The data used in bibliomining will be collected through the multimedia stations processed by the project team in a network of 15 computers.

For the dissemination and evaluation of the proposed indicators there will acquired a digital type machine XEROX 4110. The typed products on digital copying support and media effects have been accomplished with the copying/ printing 4001 machine from XEROX products. They offer an exceptional productivity, reliability and superior flexibility. Media effects being placed at the user's disposal, he is offered the flexibility to create a wide range of printing applications and finite products, such as: brochures, reports, textbooks etc.

After the experimental verification, all indicators will be proposed for accreditation and validation.

The control or the efficacy will be approved after the verification of the impact in the academic environment, library management, economic environment through the sociologic studies and methods.

Motivation of the Proposed Laboratory

University Library constitutes a part of the national system of libraries, which according to he Law of the libraries no. 334/2002, art. 9, paragraph (1), makes an integrant part of the national informational system according to this article of law, the national library system aims at the *coordinated accomplishment of the national and international library loan,* ***systemic management of the information,*** *accomplishment of the national Catalogue under partition and of the virtual national Library*. In this context, the library, standing for an information centre, has to conceive systems and methods for automatic indexation, search and retrieval of the information. The library has to be oriented towards recordings of digital information which allow mutual information exchanges and their efficient enhancement.

In the framework of the information and documentation services, there has emerged the necessity to dispose of performant indicators that should assess the quality and efficacy of the information comprised in data bases of the activity carried out. The standard SR ISO 11620:2002 – Information and documentation. Performance indicators for libraries – settles a number of assessment indicators for the evaluation of the library activity results, stating for every indicator the objective, the field of application, the definition of the indicator, the methodology, the interpretation of the data and the factors which influence the indicator, the correlated sources and indicators.

The project is based upon an attentive study of the standard SR ISO 11620:2002 which, in the last but one paragraph of the Introduction chapter, page 3, specifies: *There are certain library activities and services for which – during the period this standard was being elaborated – there was noticed the absence of* ***accredited and well documented indicators****. This fact is valid for the* ***information services****, for the users' training and for the* ***automated services*** *in general. (…).* ***The library and informational community is required to conceive the mechanisms necessary for the elaboration of adequate indicators for these aspects, paying a priority attention to this problem.***

The project intends to realize the laboratory LERISC laboratory with responsibilities at a national level in elaborating **informational models** used in the information and documentation system, which aims at the following essential aspects of the information activity: analysis and representation of the information; organization and depositing within the memory of the information; putting in standard format of the information; on-line search of the information; optimization of the search for information; testing of the interrogation interfaces and of the interrogation instruments; proposition of the assistance instruments of the search for information; accomplishment of the complex and evolutive search strategies; definition of an answer policy to the requirements of search

for information; elaboration of the methods of evaluation of the systems of search for information; identification and validation of the information sources; improvement of the functioning and administration of the information and documentation services; justification of the budgetary allowances at the disposal of the information and documentation services; evaluation of the services with direct access to the information; integration of the elements of added value; evaluation of the qualitative indicators of the provided service.

The definition of the scientometric indicators with respect to the impact of scientific information upon the users' information requirements results from the activities carried out within the information and documentation services, out of whom we can enumerate: the number of the information demands and of the accessed publications; distribution upon fields of scientific interest of the publications and of the information; the necessary of information for different disciplinary and interdisciplinary fields; orientation upon sets of themes in different fields of specialization; quality of data bases and of information sources; influence of the economic and social factors upon the systems of scientific research; the time required for processing the information demand.

In the context of the management of the information from the data bases, there can be established the deficiencies and the necessary of information for different disciplinary and interdisciplinary fields, there can be assured the quality of the development of the data bases and there can be created data bases upon scientific specialization.

LERISC Laboratory proposed for coming into being:

- is vouched for and has the possibility to educate in the spirit of the European norms and to create specialists in the field of certification of products and services;
- will pursue with coherence and consequence the process of harmonization with the European directives and norms;
- has in view a valid and credible system upon which the evaluation of the products credibility should be based;

In the activity of the information centers, in the field approached by the project, the efficacy aims at actions such as systemic management of the information. It is consequently necessary to establish the system of indicators, accredited and well documented, for the evaluation of the information management which should meet the criteria and uses SR ISO 11620 and the European community requirements.

The risk in the activity of LERISC laboratory refers to the probability of non-observance of the pre-established goals in terms of **performance** (non-accomplishment of the quality standards), **program** (non-compliance with the execution delay), and cost (exceeding of the budget).

The management of the risk stands for a cyclic process, with several distinct phases: identification of the risk, analysis of the risk and reaction to the risk. The managerial team takes into consideration both internal and external risks. The internal risks represent risks that the managerial teal may control or influence, while the external risks are beyond its control.

This way there are drawn up control lists which comprise potential sources of internal risk: expected results, staff, modifications of the goals, errors and omissions in design and execution, assessments of the costs and of the execution delays etc.; Moreover, the staff of the laboratory is invited to a formal meeting of identification of the risks. An efficient communication is one of the best sources of risk diminution and execution.

As for the external risk, there are identified the risks imposed from the exterior (through legislation, changes in economics, technology, relation with the trade-unions) by appointing a person who should take part in the meetings of the professional associations, in conferences, and who should run through the publications in the field.

The hypotheses and the risks which may affect the viability of the project are:

- accentuated modification of the currency;
- delays in the payment of the services and in the acquisition of the equipment;
- young employees with small wages may leave the organization;
- the probability of non-compliance with the pre-established goals in terms of performance.

LERISC – the laboratory we propose, which will become a structure of advanced tests for the manipulation of the sets of data for relevant information in the fields of the information science, will comply with the existing standards at an international level and those adopted at a national level.

The beneficiaries of LERISC activity will be: University libraries, component parts of the didactic process and of the national system of scientific research; Universities, by the improvement of the training for research and by the visibility of research and access to relevant information; Centers of information-documentation and research; All the users of the library; Researchers and professorial staff in all the fields and especially those in the information science; Managers of the info-documentary systems; Doctoral students and Masteral students in the Universities, for the finalization of their doctoral theses and of their dissertations; the Economic environment through the dissemination of the results by sociologic methods.

Within LERISC there will be determined the necessary mechanisms for the elaboration of accredited indicators, of use in the improvement of library

collection management, in the fluidization and filtration of the informational content and in the retrieval of the relevant information through a model based upon systems of interrogation of the users. By LERISC there will be paved the way for the knowledge society.

Technical, economic and social impact

Within modern society information mean power. And power is knowledge. Within contemporary society not only science fields, but also human being are objects of knowledge. This project-program aims at producing effect both upon human resources and upon fields of research, economic or technical. The impact upon the human resources materializes in the creation of new posts, in the creation of new specialists, of a new job.

The impact upon education and research materializes by the opportunities for education and advanced learning. The most recent discoveries, the latest research will be placed at the disposal of the researchers, of the students, of the masteral students. Within this laboratory of data mining and information mining there will be processed the required information, there will be exploited data bases in the shortest delay.

Moreover, by means of the technology and of the created system, we place ourselves at the disposal of the economic-scientific-technical environment. In a world of speed, of information avalanche, of the need of continual evaluation of the development market in any field, we must find solutions in no time. This way, it will be possible to carry out complex evaluations, statistics, to extract and index, by computing, data from the fields of interest for a time economy and for a smaller effort of administration of the great quantity of retrieved documents. This is the reason why the group of professionals who initiate this project, librarians, specialists in information and documentation, sociologists, linguists and engineers, lay the bases of a system which, besides being purely technical, is also human, and addresses people, those who need information. They will train the best infomanagers and archeologists of the web space and will form together a complex team of informational development. Specialists in information and documentation contribute essentially to the selection of the necessary information; Sociologists identify the typologies of persons and establish the categories of users according to their needs and identify the information demands of the academic and economic environment; Linguists interpret the content of the document, extract the syntactic and semantic information from the text of the document and use the information in order to harmonize them with the information necessary to the user; Engineers create the technology so that all specialists' activity should interpenetrate on the waves of the new

communication technologies; Computer scientists design and improve the logistic support, the soft for the exploitation and evaluation of the information.

TARGETS

Glossary of terms related to Bibliomining

These definitions are presented only in the context of this tutorial. Many of these terms have other definitons or more refined meanings not represented here.

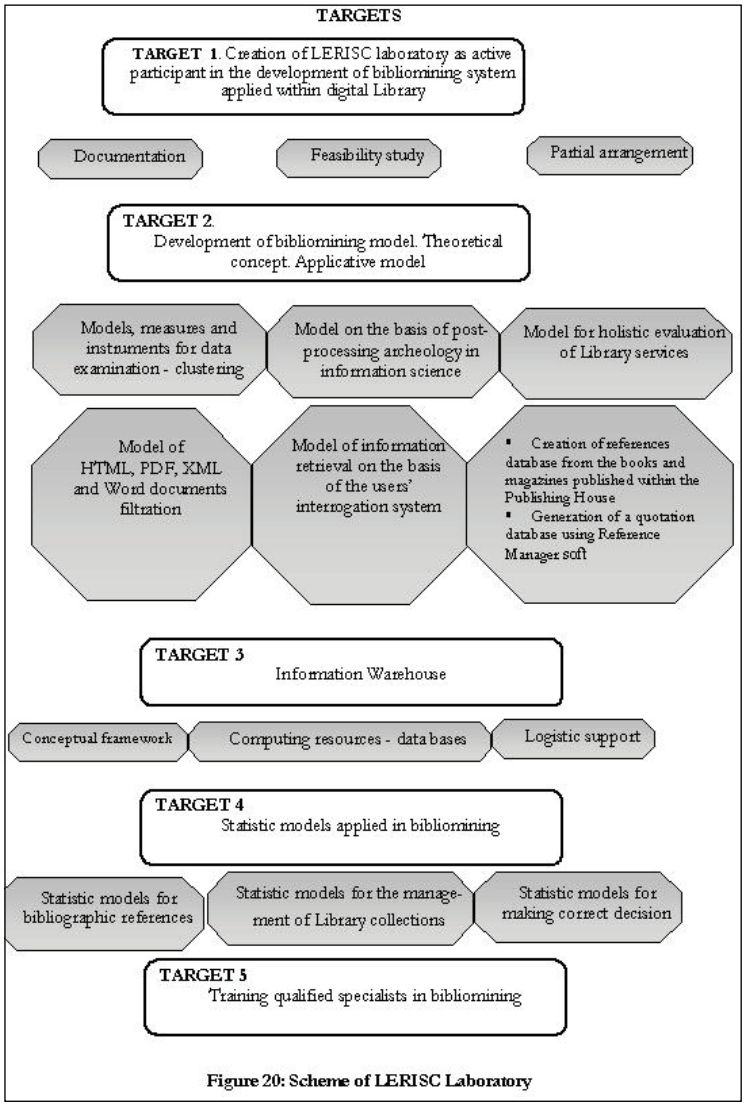


Figure 20: Scheme of LERISC Laboratory

AGGREGATION - An aggregation is used here to mean some way of combining underlying measures to create a single measure. This could be through adding a set of values (number of visits in the last year), dividing one measure by another (average items examined per search), or some other way of combining a number of underlying measures. Typically, this is done for large numbers of library users in order to provide a single measure for evaluation. The disadvantage of creating these aggregations over all library users is that it masks underlying patterns between user communities.

BIAS - A bias occurs when decisions are made regarding data collection, aggregation, data cleaning, or other steps in the bibliomining process that create a sample that is not representative of the overall population. One example is a selection bias, where individuals targeted for a user survey are not randomly selected from the population, but rather are selected because they were visiting a library or were part of some subgroup accessible to the reseacher. Another example is discarding records that have a missing value for a specific measure in order to calculate an aggregation. Few studies have no bias, but one goal of research is to reduce or control for bias.

BIBLIOMETRICS - Bibliometrics is the study of patterns in formal scholarly communication. Common items examined for bibliometrics include citations between works, patterns of authorship, journals of publication, and words used in documents. It is related to informetrics, which is the examination of patterns in a larger body of communication, and webometrics, which explores patterns in Web communication.

BIBLIOMINING - Bibliomining is the combination of data warehousing, data mining, bibliometrics, statistics, and reporting tools used to extract patterns of behavior-based artifacts from library systems. The name comes from the combination of the terms "bibliometrics" and "data mining" as bibliomining combines patterns from the creation process and patterns in the use of library services.

CORRELATION - Correlation is an indication of the relationship between two measures, and is usually measured between -1 and 1. A high positive correlation (closer to 1) means that when one measure goes up, another one goes up. A negative correlation (closer to -1) means that when one goes up, the other goes down. A correlation near 0 means that there's little observable relationship between the two measures. For example, the number of times someone visits the library should be positively correlated with the number of items that the person checks out. Correlation is useful in prediction and is the basis for many statistical procedures.

DATA MINING - Data mining is the exploration of a large cleaned data set using statistical and artificial intelligence methods. Data mining can be

directed (with a specific topic area or goal) or undirected (with no specific goal). The purpose of data mining is to discover patterns that are novel, meaningful, and actionable. In order to determine appropriate patterns, a domain expert needs to be involved with the data mining process.

DATA WAREHOUSE - A data warehouse is a secondary copy of operational data reformatted for analysis and reporting. Typically, the data warehouse will connect data from different parts of an operational system. In addition, the data warehouse can be a place for external data to be connected to operational data. Creating a data warehouse allows the time spent gathering and cleaning data for one project to be easily used in other projects.

DEMOGRAPHIC SURROGATE - A demographic surrogate is a set of variables used to replace the personally identifiable information about a patron. The demographic variables selected should be useful to the library in decision-making and justification. The result is that these surrogates can help the librarian to better understand different communities who use library services.

DOMAIN EXPERT - In the setting of a particular libary, the domain expert is someone who has worked in that library for a significant amount of time. The patterns found with data mining should correspond to the observations of the domain expert.

EVIDENCE-BASED LIBRARIANSHIP - Evidence-Based Librarianship (EBL) was originally developed from concepts used in Evidence-Based Medicine. In traditional EBL, the librarian starts by gathering research pieces that address a particular topic. The results from these pieces are then combined to create Evidence, which is then used to resolve the decision-making need. Bibliomining concepts can be used to create a different type of Evidence by collecting data from many different libraries in the same standard format. This data warehouse can then be tapped as needed for Evidence.

ONLINE ANALYTICAL PROCESSING (OLAP) - An Online Analytical Processing tool allows the exploration of a dataset through interactive manipulation of different variables over selected time periods. The resulting report can be refined through the expansion or collapse of variable categories. In regard to bibliomining, an OLAP tool gives a staff member with little training the ability to explore the library's data warehouse as needed.

OUTCOME-BASED EVALUATION - The goal of outcome-based evaluation (OBE) is to demonstrate the impact of library services. OBE starts by the selection of measurable outcomes that the library should bring about. Measures are then selected and gathered based upon these outcomes. The measures are then evaluated to determine if an impact has been made. The result is a guided evaluation plan that allows the librarians to understand what types of impact their library services are having.

OUTLIER - An outlier is a data point that is either significantly higher or lower than the bulk of the population. Many statistical methods are affected by outliers. For example, an average salary can be skewed by a small number of outliers with very high salaries. The resulting average would not provide a reasonable understanding of the desired measure.

STANDARD DEVIATION - The standard deviation of a measure indicates how spread out the individual data points are. A small standard deviation (when compared to the average) indicates that most of the individual data points are close to the average. A large standard deviation means that the average may not be very useful in understanding the data set.

STOVEPIPING - Many operational library systems keep the data for different parts of the operation separate. For example, the Interlibrary Loan data, Circulation data, and E-Resource use data usually live in three different systems. This concept, called stovepiping, makes it difficult to create library studies that look at the use of material across different systems. A data warehouse is a place where data from these different systems can be matched and formatted for analysis.

VISUALIZATION - Visualization techniques take a complex data set and attempt to make it more understandable by using graphics to display components of the data set. Graphs and charts allow for the identification of trends and comparisons of many different data points. Graphical webs showing connections between works can be used to demonstrate links or citations between works or authors. Since visualization simplifies a complex data set, it also can mask information or mislead the viewer in their interpretation of the data.

REFERENCES

1. Agent Technology: Foundations, Applications and Markets. Edited by Jennings, N.R.; Wooldridge, M.J. Berlin: Springer Verlag, 1998.
2. American Civil Liberties Union. (n.d.). USA PATRIOT act. Retrieved May 18, 2004 from http://www.aclu.org/SafeandFree/SafeandFree.cfm?ID=12126&c=207
3. BĂLĂIȚĂ, Rareş, HULEA, Mircea, OLARIU Cristian. Data Mining. Inteligență artificială – Referat, Iaşi: Universitatea Tehnică "Gh. Asachi"Iaşi, Facultatea de automatică şi calculatoare, 2002/2003
4. BALL, Philip. One-page book on horizon, Nature - Science update (on line), Tuesday 24 April 2001 ('Paper could soon be obsolete. Researchers from Bell Laboratories in New Jersey and the E Ink Corporation in Cambridge, Massachusetts, have found an economical way to make sheets of flexible plastic that can be electronically activated to display different pages of black and white lettering. A single sheet of this 'electronic paper' could, in principle, contain an entire book. Information, held in some portable memory device such as a CD player, could be fedonto the plastic page to appear as black and white characters or images. Each page of information could be called up in sequence, whereupon the 'electronic ink' on the sheet rearranges itself into a new configuration.')
5. BANERJEE, K. Is data mining right for your library? Computers in Libraries, 1998, 18(10), 28-31.
6. BARABÁSI, A. Linked. New York: Penguin Group, 2003
7. BERRY, J.; LINOFF, G. Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management. 2nd edition. Indianapolis, IN: Wiley, 2004
8. BOLLEN, J.; LUCE, R.; VEMULAPALLI, S.; XU, W. Usage analysis for the indentification of research trends in digital libraries. D-Lib Magazine, 2003, 9(5). Available online at http://www.dlib.org/dlib/may03/bollen/05bollen.html.
9. BORGMAN, C. L.; FURNER, J. Scholarly communication and bibliometrics. In B. Cronin (Ed.), Annual Review of Information Science and Technology (Vol. 36, pp. 3-72). Medford, NJ: Information Today, 2002.
10. BORGMAN, C. L. (Ed.) Scholarly Communication and Bibliometrics, Newbury Park, CA: Sage Publications, Inc., 1990.
11. BÖRNER, K.; CHEN, C.M.; BOYACK, K.W. Visualizing knowledge domains. Annual Review of Information Science and Technology 37, 2003, 179-255.
12. BOULTON, Richard E.S; LIBERT, Barry D.; SAMEK, Steve M. Cracking the Value Code. How Successful Businesses Are Creating Wealth. In the New Economy, Harper Business, New York, 2000.
13. BRADSHAW, J. Ed., Software Agents, Menlo Park, CA, AAAI Press, 1997.

14. BROWN, John Seely; DUGUID, Paul. Organizing knowledge. In: California Management Review, 40(3): 90-111, 1998 Spring.
15. BUCKLAND, M. Five grand challenges for library research. In: Library Trends, 2003, 51(4), p. 675-686.
16. CANTU-PAZ, E.; CHANDRIKA, Kamath. On the Use of Evolutionary Algorithms. In: Data Mining; http://www.llnl.gov/CASC/sapphire/pubs/heuristic.pdf
17. CHAUDHURI, S.; DAYAL, U. An overview of data warehousing and OLAP technology. In: SIGMOD Record, 1997, 26(1), 65-74.
18. CHEESEMAN, P.; J. STUTZ. Bayesian Classification: Theory and results. Eds. Fayyad, U., Piatetsky-Shapiro, G, Smyth, P. and Uthurusamy, R. In: Advances in Knowledge Discovery and Data Mining, Cambridge, MA: AAAI/MIT Press, 1996.
19. CHEN, C. Visualising semantic spaces and author co-citation networks in digital libraries. In: Information Processing and Management 35(3), 1999, .p.401-420.
20. CHIP, Elliott. Building the Wireless Internet. In: IEEE Spectrum, January 2001, p. 14-16.
21. COLLIER, K. B.; CAREY, E.; GRUSY, C.; MARJANIEMI, D; SAUTTER, D. A Perspective on data Mining, Northern Arizona University, July 1998, insight.nau.edu/downloads/DM%20Perspective%20v2.pdf
22. Commission of the European Communities, Europe 2002 – Impact and Priorities, A communication to the Spring European Council in Stockholm, 23- 24 March 2001, COM (2001) 140 final, Brussels, 13.3.2001.
23. Competing visions. In "Progress Toward Digital Libraries", eds. Christine L. Borgman, Gary Marchionini and Edward A. Fox, Special Issue, Information Processing & Management, 35(3), May, 1999, p. 227-244,
24. DAN Nica, Guvern, Cetățean, Societate informațională, Editura SEMNE, 2001.
25. DAVENPORT, Thomas D.; PRUSAK, Laurence. Working knowledge: How Organizations Manage What They Know, Harvard Buisiness School Press, 1998.
26. DRĂGĂNESCU, M. Procesarea mentală a informației, cap.5: Psihologia agenților inteligenți. In Memoriile Secțiilor științifice, Academia Română, Seria IV, Tomul XX, 1997, p. 270-272,
27. DRĂGĂNESCU, M. De la societatea Informațională la societatea cunoaă terii, Editura Tehnică, Bucureşti, 1991, p.237 www.ici.ro/cărți.html
28. DRĂGĂNESCU, Mihai. Conştiința, frontieră a ştiinței, frontieră a omenirii, comunicare la seziunea Comitetului Român pentru Istoria şi Filosofia Stiinței şi Tehnicii, Academia Română, 18 octombrie 2000. In: Revista de filosofie, XLVII, Nr.1-2, 2000, p. 15-22.
29. DRĂGĂNESCU, Mihai. Discurs la Consfatuirea Preşedintelui României cu manageri si specialişti din domeniul Tehnologiei Informației, Bucureşti,

Palatul Cotroceni, 12 aprilie 2001.
30. DRĂGĂNESCU, Mihai. Societatea conştiinţei – cuvânt la Fundaţia Henri Coandă, Atheneul Român, 30 iunie 2001.
31. DRĂGĂNESCU, Mihai. Societatea cunoaşterii şi cartea electronică, comunicare. Simpozionul Cartea Electronică, Academia Română, 21 iunie 2001.
32. DRĂGĂNESCU, Mihai. The Economy of the Knowledge Society – intervenţie la Conferinţa Noua economie - o şansă pentru România, Bucuresti, Hotel Marriott, 20 martie 2001.
33. DRĂGĂNESCU, Mihai. Conştiinţa, frontieră a ştiinţei, frontieră a omenirii, comunicare la seziunea CRIFST, Academia Română. 18 oct. 2000. In: Revista de filosofie, XLVII, Nr. 1-2, 2000, p.15-22.
34. DRĂGĂNESCU, Mihai. www.racai.ro/~dragam.
35. DREXLER, K.Eric. Nanosystems: molecular machinery, manufacturing and computation. New York, NY: John Wiley, 1992.
36. DRUCKER, Peter F. Beyond the Information Revolution, The Atlantic Monthly, Digital Edition, http://www.theatlantic.com/issues/99oct/9910drucker3.htm
37. Data Mining, Turn computer loose on your data and you don't know what they'll come up with--that's the whole point, coordonator E.X. de Jesus, In: BYTE, October, 1995. http://www.byte.com/art9510/sec8/sec8.html
38. EIRINAKI, M.; VAZIRGIANNIS, M. Web mining for web personalization. ACM Transactions on Internet Technology, 2003, 3(1), p.1-27.
39. EROIL Gelenbe; ŞEREF Esin; XU Yhiguang, Simulation with Learning Agents, p.148-157.
40. European Commission Enterprise DG - European Observatory on Intangible Assets, Policy Trends in Intangible Assets, 8 Nov. 2000 (prin Internet, http://www.eu-intangibles.net).
41. European Commission, Community research, ISTAG, Scenarios for Ambient Intelligence in 2010, Final Report, February 2001, Seville.
42. FAYYAD, U.M.; PIATETSKI-SHAPIRO, G.; SMYTH P.; UTHURUSAMY R. Advences in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996.
43. FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Data bases. In: AI Magazine, Fall 1996a.
44. FAYYAD, U. Editorial. Data Mining and Knowledge Discovery, 1997.
45. FAYYAD, USAMA; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery: An Overview: In: Advances in Knowledge Discovery and Data Mining, AAAI/MIT Press, 1996b.
46. FAYYAD, USAMA; PIATETSKY-SHAPIRO, G.; SMYTH, P. Knowledge Discovery and Data Mining: Towards a Unifying Framework. In: KDD-96 Conference Proceedings. Ed. E. Simoudis, J. Han, and U. Fayyad, AAAI

Press, 1996c.
47. FAYYAD, Usana; UTHUTASAMY, Ramasamy. Data Mining and Knowledge Discovering in Data bases. In: Comm. of ACM, November 1996, vol.39, nr.11, p.25-26.
48. FRAWLEY, W.; PIATETSKY-SHAPIRO, G.; MATHEUS; C. Knowledge Discovery in Data bases: An Overview. In: AI Magazine, Fall 1992, pp 213-228.
49. GARDNER, Claudia. IBM Data Mining Technology, 11.04.96. http://booksrv2.raleigh.ibm.com/cgi-bin/book.mgo.cmd/BOOKS/datamine sau http://www.almaden.ibm.com/cgi-bin/stss/data-mi
50. GEPPERT, Linda. The New Chips on the Block. In: IEEE Spectrum, January 2001, p. 66-68.
51. GEYER-SCHULZ, A.; HAHSLER, M.; NEUMANN, A.; THEDE, A. (2003). An integration strategy for distributed recommender services in legacy library systems. Editors: M. Schader, W. Gaul, and M. Vichi, In: Between Data Science and Applied Data Analysis, Proceedings of the 26th Annual Conference of the Gesellschaft für Klassifikation e.V., University of Mannheim, July 22-24, 2002, Studies in Classification, Data Analysis, and Knowledge Organization, pages 412-420. Springer-Verlag, July 2003.
52. GRAYSON, C. et al. Mining your hidden resources. In: Across the Board, 35(4): 23-28, April 1998.
53. GREBNER, Ceryl. Excavate Your Data. In: Datamation, May 1, 1996, p.40-41.
54. GURAN, Marius. Societatea informațională; propuneri pentru o strategie națională – studiu, aprilie 2001, prezentat la ședința comună a Secției pentru știința și Tehnologia Informației și Forumului pentru Societatea Informațională ale Academiei Române.
55. HAN, J.; SIMOUDIS, E. Proceedings of the First International Conference on Knowledge Discovery and Data Mining (KDD-95), AAAI Prress, 1995, http://www.aaai.org/Publications/Catalog/KDD și http://www-aig.jpl.nasa.gov/KDD95
56. HAND, D.; MANNILA, H; SMYTH, P. Principles of Data Mining. MIT Press, Cambridge, MA, 2001
57. HOPKINS, John F., FISHWICK, Paul A. A Three-Dimensional Human Agent Metaphor for Modelling and Simulation, p.131-147.
58. http://en.wikipedia.org/wiki/Data Mining (accesat în 10.01.2006)
59. http://info.gte.com/~kdd/index.html (accesat în 14.02.2006)
60. http://www.angoss.com/ (accesat în 28.01.2006)
61. http://www.anvilinformatics.com/ (accesat în 22.02.2006)
62. http://www.cio.com/archive/051598_mining.html (accesat în 10.03.2006)
63. http://www.cognos.com/ (accesat în 23.02.2006)
64. http://www.cs.bham.ac.uk/~ano/TheDataMine.html (accesat în 19.02.2006)

65. http://www.informatik.univ-trier.de/~ley/journals (accesat în 17.01.2006)
66. http://www.microsoft.com/sql/productinfo/datamine.htm (accesat în 18.02.2006)
67. http://www.oracle.com/ip/analyze/warehouse/datamining/ (accesat în 15.01.2006)
68. http://www.pilotsw.com/r_and_t/whatpaper/datamine/ (accesat în 24.02.2006)
69. http://www.scitegic.com/products_services/pipeline_pilot.htm (accesat în 28.02.2006)
70. http://www.tcs.hut.fi/~helger/crypto/link/data_mining (accesat în 05.02.2006)
71. http://www.uts.edu.au/fac/hss/Departments/DIS/km/introduct.htm#Char (accesat în 17.02.2006)
http://www.uts.edu.au/fac/hss/Departments/DIS/km/Resources/KMIntro/index.htm (accesat în 11.01.2006)
72. http://www3.shore.net/~kht/text/dmwhite/dmwhite.htm (accesat în 12.03.2006)
73. http://www-4.ibm.com/software/data/iminer/fordata/about.html (accesat în 13.02.2006)
74. http://www.megaputer.com/dm/index.php3 (accesat în 09.01.2006)
75. http://www.std.com/CreativeData/credata/termin.html (accesat în 15.01.2006)
76. http://www.racai.ro/books/doe (accesat în 30.01.2006)
77. IANCU, Ştefan. Managementul cunoştinţelor, Univers ingineresc, XII, 2001, Nr. 3, 4, 6,7, 8, 10.
78. IEEE Transaction on Knowledge and Data Engeneering, vol.8, nr.6, December 1996, http://www.computer.org/pubs/fkde/tok.html, IEEE Expert, October, 1996, http://www.computer.org/pubs/expert/feature.html, Communications of ACM, vol.39, 11, November 1996
79. IEEE Transactions on Neural Networks, vol.10, 1999; vol.11, 2000; vol.12, 2001.
80. Information Institute of Syracuse (2004). QABuilder. Retrieved May 21, 2004 from http://vrd.org/qabuilder.shtml
81. Interoperability, Scaling, and the Digital Libraries Research Agenda, report of the 1995 IITA DL Workshop
82. IONIŢĂ, Angela. Asupra termenului de minerit de date (Data Mining), Institutul de Cercetare pentru Inteligenţă Artificială, Academia Română, http://www.ici.ro/ici/revista
83. ISO/IEC 13249-6 SQL/MM Data Mining. Edited by Friedemann Schwenkreis, IBM Deutschland Entwicklung GmbH http://research.microsoft.com/~jamesrh/hpts2001/submissions/FriedemannSchwenkreis.htm
84. JACKSON, J.E. An User's Guide to Principial Components, New York,

NY: John Wiley, 1991.
85. JOHNSON, M. Archeological Theory: An Introduction. Oxford: Blackwell, 1999.
86. KANTE, Rosabeth Moss. Evolve! Succeding in the Digital Culture of Tomorrow, Harvard Business School Press, Boston, Massachusetts, e-Book MSReader edition, 2001.
87. KAO, S.; CHANG, H.; LIN, C. Decision support for the academic library acquisition budget allocation via circulation database mining. In: Information Processing &Management, 2003, 39(1), 133-148.
88. KDD-97, The Third International Conference on Knowledge Discovery and Data Mining, August 14-17, 1997
89. KOGAN, A.; SUDIT; F.; Vasarhelyi M. The Internet Guide for Accountatnts: a living book, Prentice Hall, 1997. sau U.Fayyad, G.Piatetsky-Shapiro, P.Smyth, The KDD Process for Extracting Useful Knowledge from Volumes of Data, Comm. of AC M, 39, 11, p.27-34.
90. KOSTOFF, R., del Río; HUMENIK, J.; GARCÍA, E.; RAMÍREZ, A. Citation Mining: Integrating Text Mining and Bibliometrics for Research User Profiling. In: Journal of the American Society for Information Science and Technology, 2001, 52(13), 1148-1156.
91. KOULOPOULOS, Thomas M.; SPINELLO, Richard A.; WAYNE, Toms. Corporate Instinct: Building a Knowing Enterprise for the 21st Century, Van Nostrand Reinhold, 1997
92. KRIVADA, Ceryl D.; Unearthing Underground Story Data, Cover Story, May, 1996, http://www.lmag.com/9605mine.html
93. LEONARD-BARTON, Dorothy. Wellsprings of Knowledge: Building and Sustaining the Sources of Innovation, Harvard Business School Press, 1955.
94. LIMPAA, H. Privacy-preserving Data Mining. Retrieved May 18, 2004
95. LOGAN, Brian; THEODOROPOULOS, Georgios. The Distributed Simulation of Multiagent Systems, in Special issue on agents in modelling and simulation: exploiting the metaphor. Eds. A.M. Uhrmacher, P.A. Fishwick, B.P. Zeigler, In: Proceedings of the IEEE, february 2001, p. 174-185.
96. LUCEY, T. Management Information Systems, DP Publications Ltd., Ed.6, 1991.
97. MADHAVAN, Ravindranath; GROVER, Rajiv. From embedded knowledge to embodied knowledge: New product development as knowledge management. In: Journal of Marketing. 62(4), 1-12, 1998 Oct..
98. MAFRUZ ZAMAN, ASHRAFI D.; TANIAR, K.; SMITH, A. A Data Mining Architecture for Clustered Environments, Publisher: Springer-Verlag Heidelberg, ISSN: 0302-9743
99. MARCHIONINI, Gary; FOX, EDWARD A. “Progress toward digital libraries: augmentation through integration”, guest editors’ introduction to

"Progress Toward Digital Libraries". Eds. Gary Marchionini and Edward A. Fox, Special Issue, Information Processing & Management, 35(3), May 1999, pp. 219-225.
100. MARCHIONINI, Gary; FOX, EDWARD A. "Progress toward digital libraries: augmentation through integration". In: Information Processing & Management, 35(3):219-225, May 1999
101. MARSHALL, Lucy. Facilitating knowledge management and knowledge sharing: New opportunities for information professionals, Online. 21(5): 92-98. 1997 Sep/Oct..
102. MCCLURE, C. Increasing the usefulness of research for library managers: Propositions, issues, and strategies. Library Trends, 1989, 38(2), 280-294.
103. MCCLURE, C.; LANKES, R. D.; GROSS, M.; CHOLTCO-DEVLIN, B. (2002). Statistics, Measures, and Quality Standards for Assessing Digital Library Services: Guidelines and Procedures. Syracuse, NY: ERIC Clearinghouse on Information & Technology. Retrieved May 21, 2004 from http://quartz.syr.edu/quality/
104. MENZIES, T.; HU, Y. Data Mining For Very Busy People. In: IEEE Computer, October 2003, pp 18-25.
105. MERKLE, Ralph C. Nanotechnology: What Will It Mean? In: IEEE Spectrum, January 2001, p.19-21.
106. META Group Application Development Strategies. Data Mining For Data Warehouse: Uncovering for Hidden Patterns, 13.07.1995 and Pilot Software Group, Dat Mining - Discovering Hidden Values in Your data Warehouse, Whitepaper
107. MICHAELS, James W. How New is the New Economy? In: Forbes, October 11, 1999, p.47.
108. MICHAIL, A. Data mining library reuse patterns in user-selected applications. 14th IEEE International Conference on Automated Software Engineering, Washington, DC. In: IEEE Computer Society, 1999, 24-33.
109. Mobile Agents '98, Edited by Rothermel, M.J.; F.Hohl, New York: Springer Verlag, 1998.
110. MOORE, Samuel K. In: IEEE Spectrum, January 2001, p.25.
111. MURPHY, D.E. Some librarians use shredder to show opposition to new F.B.I. powers. In: New York Times, 2003, April 7, pp. A12.
112. National Academy of Sciences (USA), Building a Workforce for the Information Economy, Report, National Academy Press, Washington, D.C., 2001, Internet edition on http://books.nap.edu/htm/IT_workforce
113. NICHOLSON, S.; STANTON, J. Gaining strategic advantage through bibliomining: Data mining for management decisions in corporate, special, digital, and traditional libraries. Eds. H. Nemati and C. Barko, Organizational data mining: Leveraging enterprise data resources for optimal performance (pp.247-262). Hershey, PA: Idea Group Publishing, 2003.

114. NICHOLSON, S. The bibliomining process: Data warehousing and data mining for library decision-making. In: Information Technology and Libraries, 2003, 22 (4), 146-151.
115. NICHOLSON, S. Bibliomining bibliography. The Bibliomining Information Center. Retrieved March 1, 2004a, from http://bibliomining.org
116. NICHOLSON, S. A conceptual framework for the holistic measurement and cumulative evaluation of library services. Journal of Documentation, 2004b, 60(2), 162-182.
117. NICHOLSON, S. A framework for Internet archeology: Discovering use patterns in digital library and Web–based information resources. First Monday 10(2). Retrieved March 23, 2005 from http://www.firstmonday.org/issues/issue10_2/nicholson/index.html
118. NICHOLSON, Scott. The Basis for Bibliomining: Frameworks for Bringing Together Usage-Based Data Mining and Bibliometrics through Data Warehousing in Digital Library Services. In Information Processing and Management
119. NICOLAE, Gabriel; DRAGOMIRESCU, Liviu; BARBOS, Florin; HOBEANU, Claudiu; ICRIVERZI, George. Complemente de informatica
120. NISO. (2004). Z39.7 library statistics – E-metrics data dictionary. Retrieved May 18, 2004 from http://www.niso.org/emetrics/
121. NITCHI, S.I.; AVRAM-NITCHI, R. Data Maning, o nouă eră în informatică.
122. NITCHI, S.I.; RACOVITAN, D. – Coordonatori. Bazele prelucrării informatiilor si tehnologie informationalã. Ed. Intelcredo, 1996.
123. NONAKA, Ikujiro; HIROTAKA; Takeuchi. The Knowledge Creating Company, Oxford University Press, 1995.
124. PAMBUCCIAN, Varujan. Expunere la lansarea Programului Societatea Informațională – Societatea cunoaşterii (SI-SC), Academia Română, 10 aprilie 2001.
125. PATEL, L.; PERRY, S.; TAYLOR, D.; BROWN, A.; TIKE, S. The Future of Data Mining http://www.aston.ac.uk/~golderpa/CS342/grouppages/dssg9/fut.html
126. PEIRCE, C. S.: Deduction, induction and hypothesis. În: Popular Science Monthly, 13, 1878.
127. Pilot Software. An Introduction to OLAP, http://www.pilotsw.com/r_and_t/whtpaper sau Creative Data, Data Warehouse Terminology
128. PITAC Report, "Biblioteci digitale: Acces universal la cunoaşterea umană", 2001, p. 3
129. PRODI, Romano. Speech at the German Confederation of Trade Unions' (DGB) Federal Presidium, Brussels, 7 November 2000, apud Cordis

Focus, European Commision, 20 November 2000, p.1 and 3.
130. Project Counter. Counter – Counting Online Usage of Networked Electronic Resources, Retrieved May 18, 2004 from http://www.projectcounter.org
131. ROONEY, David; Mandeville, Thomas. The Knowing Nation: A Framework for Public Policy in a Post-industrial Knowledge Economy, In: Prometheus, 1998, 16 (4) pp. 453-467.
132. QUÉAU, Philippe - Directeur de la division de l'information et de l'informatique de l'Unesco, La necessaire définition d'un bien public mondial - A qui appartiennent les connaissances? Le monde diplomatique, Janvier 2000, pages 6 et 7. http://www.monde-diplomatique.fr/2000/01/QUEAU/13278.html
133. Rensselaer Polytechnic Institute, Data Mining and Knowledge Discovery. http://www.vpi.edu/~vanepa2/
134. RIEZEMAN, Michael J. Optical Nets Brace for Even Havier Traffic. In: IEEE Spectrum, January 2001, p. 44-46.
135. ROGERS, J. A. et al. Paper-like electronic displays: large-area rubber-stamped plastic sheets of electronics and microencapsulated electrophoretic inks. In: Proceedings of the National Academy of Sciences USA, 2001, 98, 4835–4840.
136. SANDSTROM, P.E. Scholarly communication as a socioecological system. In: Scientometrics, 2001, 51(3), 573-605.
137. SARACEVIC, T.; KANTOR, P. Studying the value of library and information services: Part 1, Establishing a theoretical framework. In: Journal of the American Society for Information Science, 1997, 48(6), 527-542.
138. SARJOUGHIAN, Hessam S.; YIEGLER, Bernard P.; HALL, Steven B. A Layered Modeling and Simulation Architecture for Agent-Based System Development, in Special issue on agents in modelling and simulation: exploiting the metaphor, eds. A.M. Uhrmacher, P.A. Fishwick, B.P. Zeigler. In: Proceedings of the IEEE, february 2001, p.201-213.
139. SCHILLER, Dan. Digital Capitalism, MIT Press, Hardcover 1999, Paperback 2000, eBook (MSReader) 2000.
140. SIMOUDI, E.; HAN, J.W.; FAYYAD; U. Procedings of the Second International Conference Discovery and Data Mining (KDD-96), AAAI Press, http://www-aig.org/Publications/Catalog/KDD/, sau http://www-aig.jpl.nasa.gov/KDD95/
141. Source Book on Digital Libraries. editor Edward A. Fox, 1993, pg. 65
142. SOUTH, S. Method and Theory in Historical Archeology. New York: Academic Press., 1977
143. SRIVASTAVA, J.; COOLEY, R.; DESHPANDE, M.; TAN, P.T. Web usage mining: Discovery and applications of usage patterns from Web data. In: SIGKDD Explorations, 2000, (1)2, 12-23.

144. STEWART, Thomas A., Intellectual Capital: The New Wealth of Organizations, Currency/Doubleday, 1997
145. STYX, Gary. Trends in Nanotechnology; Waiting for Breakthroughs, Scientific American, April 1996, p.78-83.
146. TAIPALE, K. A. Data Mining and Domestic Security: Connecting the Dots to Make Sense of Data (http://www.stlr.org/cite.cgi?volume=5&article=2), Center for Advanced Studies in Science and Technology Policy (http://www.advancedstudies.org/). 5 Colum. Sci. & Tech. L. Rev. 2 (December 2003).
147. TALPEŞ, Florin - Managerul general Softwin. Inițiativa Open e-book pe plan internațional– comunicare, Simpozionul Cartea Electronică, Academia Română, 21 iunie 2001.
148. TECUCI, Gheorghe. Building Intelligent Agents, An apprentiship Multistartegy Learning Theory, Methodology and Case Studies, Academic Press, 1998.
149. TEODORESCU, Horia-Nicolai; LAKHMI, C. Jain; KANDEL, Abraham. Hardware Implementation of Intelligent Systems, Physica-Verlag, Springer-Verlag Company, Heidelberg, 2001.
150. The Corporate KDD Bookmark, http://www.cs.su.oz.au/~thierry/ckdd.html
151. The Scope of the Digital Library, Draft Prepared by Barry M. Leiner for the DLib Working Group on Digital Library Metrics, 1998
152. THEARLING, Kurt. From Data Mining to Data Base Marketing, White Paper, 95/02, October 1995, Data Inteligent Group, Pilot Software, http://www.santafe.edu/~kurt/wp9502.html
153. WAGNER, Gerd. The AOIS Glossary. On Agent-orientation in Information Systems.
154. WHITE, H.D.; MCCAIN, K.W. Bibliometrics. In M.E. Williams (Ed.) In: Annual Review of Information Science and Technology 24. Medford, NJ: Information Today, 1989, p.99-168.
155. WHITE, H.D.; MCCAIN, K.W. Visualizing a discipline: An author co-citation analysis of information science 1972-1995. In: Journal of the American Society of Information Science, 49(4), 1998, 327-355.
156. WILKINSON, D.; THELWALL, M.; Li, X. Exploiting hyperlinks to study academic Web use. In: Social Science Computer Review, 21(3), 2003, 340-351.
157. WILSON, Linda F. and others, A Framework for Linking Distributed Simulations Using Software Agents, în Special issue on agents in modelling and simulation: exploiting the metaphor, eds. A.M. Uhrmacher, P.A. Fishwick, B.P. Zeigler. In: Proceedings of the IEEE, february 2001, p. 186-199.
158. WURSTER, Thomas S.; EVANS; Philip. Blown to bits. How the new economics of information transforms strategy, Harvard Business School

Press, Boston, Massachusetts, Hardcover, 2000. Microsoft Reader edition (eBook), 2000.
159. WURSTER, Thomas S.; EVANS, Philip. Strategy and the New Economics of Information, Harvard Business Review, Sept.-Oct. 1997.
160. ZUCCA, J. Traces in the clickstream: Early work on a management information repository at the University of Pennsylvania. In: Information Technology and Libraries, 22(4), 2003, 175-179.
161. * * *: IBM's Data Mining Technology - White Paper data Management Solutions, April, 1996 (http://cgi.di.uoa.gr/~rouvas/research/ibm/IBM-datamine.html)
162. * * *: NASA Workshop on Issues in the Application of Data Mining to Scientific Data, October 19-21, 1999, University of Alabama in Huntsville, Huntsville, Alabama, www.cs.uah.edu/~thinke/NASA_Mining/DMFinalReport.pdf
163. * * *: Wikipedia article "Data mining", http://en.wikipedia.org/wiki/Data Mining
164. *** Conferința Noua Economie – O şansă pentru România, organizată de Arthur Andersen şi Oracle, sub patronajul Primului Ministru al României, Adrian Năstase, în colaborare cu Ministerul Comunicațiilor şi Tehnologiei Informației _i Ministerul Industriei şți Resurselor, Bucureşti, Hotel Marriott, 20 martie 2001.
165. *** eEurope+ - Action Plan, Europe 2003, A co-operative effort to implement the Information Society in Europe, Action Plan prepared by the Candidate Countries with the assistance of the European Commission, June 2001.
166. *** Europe seen leading the way for Internet standard, ştire CNN, Web posted at May 17, 2001.
167. *** European HLEG report on 'The intangible economy impact and policy issues', after CORDIS focus, 18 december 2000, p.8.
168. *** Göteborg Statement Summit of the European Union and the United States of America, 14 June 2001.
169. *** Inovation and Technology Transfer, Special Edition, The European Commission, Innovation/SMEs Programme, November 2000, p.21.
170. *** MIT to make nearly all course materials available free on the World Wide Web. Unprecedented step challenges 'privatization of knowledge', News Realease, April 4, 2001, http:// web.mit.edu/newsoffice/www.
171. *** Nano: the new buzzword in science and technology, CORDIS focus, 20 November 2000, p. 4-5.
172. *** Special Issue- The Limits of Semiconductor Technology, Proceedings of the IEEE, march 2001.