Vision of Semantic Processing and the Latest Trends

Nerutė KLIGIENĖ
Institute of Mathematics and Informatics, Vilnius Gediminas Technical University, Lithuania
nerute@ktl.mii.lt

INFORUM 2004: 10th Conference on Professional Information Resources
Prague, May 25-27, 2004

Abstract. *During the last decade, the emphasis has been on finding relevant documents or content, an objective which most of today's search and browsing techniques address. During the next decade, the emphasis will shift from documents and entities to relationships - that of discovering or validating contextually relevant, meaningful and possibly complex relationships amongst the entities that documents mention and describe. This also is likely to be the next area of focus for the Semantic Web community after it makes progress in addressing the current areas of emphasis on ontology representation, semantic metadata extraction for automatic annotation, query processing and making inference.*

Relationships are fundamental to semantics - they associate meanings to words, terms and entities. Semantic Web intends to associate annotations (i.e., metadata) with all Web-accessible resources such that programs can associate "meaning with data" to interpret them, and to process (access, invoke, utilize, and analyse) them automatically, resulting in higher scalability and better productivity. More comprehensive set of relationships that may be based on information extraction, external and prior knowledge and user-defined computations is looked for. Benefits provided by e-prints, IT tools and new roles of librarians as information managers in the scientific library context are considered in this presentation.

1. Introduction

For centuries, philosophers have sought universal categories for classifying everything that exists, lexicographers have sought universal terminologies for defining everything that can be said, and librarians have sought universal headings for storing and retrieving everything that has been written.

Today emerging the concept of semantic web has enlarged the task to the level of classifying, labelling, defining, finding, integrating, and using everything on the World Wide Web, which is rapidly becoming the universal repository for all the accumulated knowledge, information, data, and garbage of humankind.

The rise of the World Wide Web has created an urgent need to define standard methods and vocabularies for describing its contents in a consistent and orderly manner. Since 1995 a number of related initiatives have arisen in what has been called a Metadata Movement. Metadata is a broad term that covers many types of "structured data about data" – from traditional resources such as library catalogues, to new forms of technical and descriptive data for Web resources.

Some of these metadata types are meant to be read by human, while others are designed to be processed directly by computers. No single type of metadata can suit every application, every type of resource, and every community of users. Rather, the broad diversity of potential metadata needs can best be met by a multiplicity of separate but functionally focused metadata packages or schemas. Schemas that are designed to cover the basic needs of users in a wide range of fields and applications are called "core" element sets. In the Web environment, the most important of these is the Dublin Core element set – a set of fifteen broad categories such as *Title*, *Creator*, *Subject*, *Publisher*, and *Date*. This elementary set was designed to be simple and intuitive enough for ordinary people to use without special training.

The Internet is a giant semiotic system. It is a massive collection of three kinds of signs: *icons*, which show the form of something; *indices*, which point to something; and *symbols*, which represent something according to some convention.

The subject of *ontology* is the study of the *categories* of things that exist or may exist in some domain. The product of such a study, called *an* ***ontology***, is a catalogue of the types of things that are assumed to exist in a domain of interest ***D*** from the perspective of a person who uses a language ***L*** for the purpose of talking about ***D***.

Philosophers often build their ontology from the top down with grand conceptions about everything in heaven and earth. Programmers, however, tend to work from the bottom up. Each of them has been called a solution to all the world's problems; and each of them has been successful in solving some of the world's problems. But none of them has achieved the ultimate goal of integrating everything around a unified schema.

2. Some Basic Notions

The terms introduced here will be exploited and discussed in the sections below.

Semantic Web

The latest attempt to integrate all the world's knowledge is the *Semantic Web*. That is useful, but the problems of syntax are almost trivial in comparison to the problems of developing a common or at least a compatible semantics for everything.
So far, its major contribution has been to propose Extensible Markup Language (XML) as the common syntax for everything. XML is a simple, very flexible text format derived from SGML (Standard Generalized Markup Language). Originally designed to meet the challenges of large-scale electronic publishing, XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere.
Surprisingly, the definition of *Semantic Web* was not found in the *Webopedia* – The Online Encyclopedia dedicated to computer technology as retrieved in April 2004.

XML

Short for ***Ex****tensible* ***M****arkup* ***L****anguage,* a specification developed by the World Wide Web Consortium (W3C). XML is a pared-down version of SGML, designed especially for Web documents. It allows designers to create their own customized tags, enabling the definition, transmission, validation, and interpretation of data between applications and between organizations.

URI

Short for ***U****niform* ***R****esource* ***I****dentifier,* the generic term for all types of names and addresses that refer to objects on the World Wide Web. A *URL* is one kind of URI.

URL

Abbreviation of ***U****niform* ***R****esource* ***L****ocator,* the global address of documents and other resources on the World Wide Web. The first part of the address indicates what protocol to use, and the second part specifies the IP address or the domain name where the resource is located. For example, the two URLs below point to two different files at the domain *pcwebopedia.com:*

ftp://www.pcwebopedia.com/stuff.exe
http://www.pcwebopedia.com/index.html

The first specifies an executable file that should be fetched using the FTP protocol; the second specifies a Web page that should be fetched using the HTTP protocol.

PURL
Short for ***P****ersistent* ***URL****,* a type of URL that acts as an intermediary for a real URL of a Web resource. When you enter a PURL in a browser, the browser sends the page request to a PURL server, which then returns the real URL of the page. PURLs are *persistent* because once a PURL is established, it never needs to change. The real address of the web page may change but the PURL remains the same.

RDF
Short for ***R****esource* ***D****escription* ***F****ramework*. RDF is a general framework for describing a Web site's metadata, or the information about the information on the site. It provides interoperability between applications that exchange machine-understandable information on the Web. RDF details information such as a sitemap, the dates of when updates were made, keywords that search engines look for and the Web page's intellectual property rights. Developed under the guidance of the World Wide Web Consortium, RDF was designed to allow developers to build search engines that relay on the metadata and to allow Internet users to share Web site information more readily. RDF relies on XML as interchange syntax, creating an ontology system for the exchange of information on the Web.

3. New Trends, New Means: Semantic Web, Open E-print Archives

Intelligent agents (IA) - software robots were invented when the World Wide Web was already shaping up to be more information than any unaided human could sift through. In the meantime, the web has continued to grow. IA was expected to forage day and night for news and information, to know all about our needs, likes and interests. It never happened. Search engines were harvesting for keywords, textual information but they were not able to detect relationships among them, distinguish what was essential and what was not.

One example of such occurrences: the web site devoted to the Resources of Lithuanian cultural heritage – the ethnological, linguistic, folklore, and archeological data base has been created and named in Lithuanian – **aruodai (Granary).**
In spite that all metadata (keywords, description, title) were made correctly, the rating of this web site in the Google was very low, displayed among 60 - 70 results, if searching as a Lithuanian cultural heritage unit, but become the No 1, if retrieved as “symbolic aruodai”. But who of normal users could guess such turn of things? Searching for “aruodai”, a digital tourist will get a list of URLs related to traditional corn-bin matters where no grains of cultural heritage available…That is curiosities of a simple web.
By the late 1990s, leading search engine of the day, Altavista, could index only 30 % of the Web, while Google is striving to match the Web’s astonishing growth. Can we really

build intelligence into the 3 or 10 billion documents that make up the Web? These robots were hard to build; some of the best information on the Web was hidden in databases that agents couldn't enter. Now the problem is attacked from the other direction. Web must be made more data-like, more amenable to computer understanding, web pages have to contain their own semantics. This leaded to the Semantic Web, based on XML – the Extensible Markup Language, instead of HTML used in a simplistic Web. The World Wide Web Consortium, shortly called W3C, spearheads this goal. The Semantic Web is just one item in W3C diverse agenda, but it is increasingly important – four interest groups are working on its technologies.

E-prints Collections and Open Archives Initiatives

In the survey of scholarly publishing from the early days to the present day Correia, Teixeira (2002) stress the emerging new publishing models: e-prints / e-scripts servers and the benefits they provide to scientific community. The international initiatives towards the creation of a global network of cross-searchable research materials outlined in that survey. In order to exploit fully the expansion in the number of e-prints repositories distributed across the Internet it was recognized that there was a need to ensure that searches could be made across different e-print archives.

The OAI – Open Archives Initiative addresses this issue; the initiative emerged from the Santa Fe Convention held in 1999. The OAI aims to create cross-searchable databases of research papers and make them freely available on the web by developing and promoting interoperability standards that will facilitate the efficient dissemination of content. Using these standards, institutions can put content on the Internet in a manner that makes individual repositories interoperable. At the center of this work is the *OAI Metadata Harvesting Protocol*. This creates the potential for interoperability between e-prints archives by enabling metadata from a number of archives to be harvested and collected together in a searchable database. The metadata harvested is in the Dublin Core format and normally includes only a few metadata elements such as author, title, subject, abstract, and date.
The eprints.org at the University of Southampton provides free software that enables any institution to install OAI-complaint archives (i.e. using the OAI metadata tags). It is designed to run centralized, discipline-based as well as distributed, institution-based archives of scholarly publications (Chan and Kirsop 2001). OAI-compliant e-prints servers provide value added facilities. They can compile statistics which show authors how many times their papers have been accessed; they can also produce an online publication list by author or by academic department. Furthermore, developing services such as the *OpCit – the Open Citation Project* aim to provide integration and navigation through citation linking and give authors citation and impact analysis of their work (Nottingham ePrints).

4. The Innovations Behind the Semantic Web

The ideas behind the Semantic Web are innovations that simply extend current Web techniques in ways that make documents more data-like, so that agents can interact with them in sophisticated ways. Those ideas are: the Uniform Resource Identifiers (URI), the Extensible Markup Language (XML), and the Resource Description Framework (RDF). The terms have been shortly introduced in this paper, the Basic Notions section, below we are considering them in more details.

First of all innovations it is based on URIs (Uniform Resource Identifiers) what is more general than URLs (such as Web address, a link to an entity on the web). URI identifies resources in general: items like human beings, corporations and bound books in a library are resources, just not “network retrievable” ones. In generic terms the Web is an information space. Human beings have a lot of mental machinery for manipulating, imagining, and finding their way in spaces. URIs are the points in that space. More exactly, URIs is the naming/addressing technology. Unlike web data formats, where HTML is an important one, but not the only one, and web protocols, where HTTP has a similar status, there is only one Web naming/addressing technology - URIs. URIs are short strings that identify resources in general: documents, images, downloadable files, services, electronic mailboxes, and other resources. They make resources available under a variety of naming schemes and access methods such as HTTP, FTP, and Internet mail addressable in the same simple way.

XML builds on the second fundamental web technique: coding elements in a document. HTML coding serves mostly to control the appearance and arrangement of the text and images on a Web page, so that only a few elements are tagged, such as <title> for an article’s title, <bold> for boldface type and <table> to begin a table. They identify document only stylistically. XML singles things out as data elements – as dates, prices, invoice numbers, and so on. It allows users to mark up any data elements whatsoever.

The Resource Description Framework, called RDF, is the third component of the Semantic Web. An RDF makes it possible to relate one URI to another. It is a sort of statement about entities, often expressing a relation between them. An RDF might express, for example, that one item is a part of another, or that a new auction bid is greater than the current high offer. Ordinary statement can’t be understood by computers, but RDF-based statements are computer-intelligible because XML provides their syntax – marks their parts of speech. RDF is a language for representing information about resources in the World Wide Web. It is particularly intended for representing metadata about Web resources, such as the title, author, and modification date of a Web page, copyright and licensing information about a Web document, or the availability schedule for some shared resource. However, by generalizing the concept of a "Web resource", RDF can also be used to represent information about things that can be *identified* on the Web, even when they *can't be directly retrieved* on the Web. RDF provides a common framework for expressing this information so it can be exchanged between applications without loss of meaning.
The notion that ties all the others together is that of ***ontology*** – a collection of related RDF statements, which together specify a variety of relationships among data elements and ways of making logical inferences among them. For example, “*syntax*”, “*semantics*”, and “*ontology*” are concepts of linguistics and philosophy but also used by the theorists in the Semantic Web community and their meanings are not changed.
An extensive list of references on ontology can be find in the works of Fensel (2001, 2002) and a list of software tools for creating the ontology’s environment can be easily completed: Apollo, LinkFactory®, OILEd, OntEdit, Ontolingua Server, OntoSaurus, OpenKnoME, Protégé-2000 SymOntoX, WebODE, WebOnto. One of them - Protégé-2000 allows the user to construct domain ontology, enter domain knowledge, and customize knowledge-acquisition forms. A platform can be extended with graphical widgets for tables, diagrams, animation components to access other knowledge-based systems embedded applications.

5. The Expansive Vision of the Next Web

The current king of the Web search world, Google, doubts the Web will ever be remade so as to be navigable by computers on their own. One of reasons is that the most Web page creators have little incentive to do detailed mark-up required by XML. But specific corners of the Web may use XML encoding and Semantic Web intelligence successfully. One of early example of the Semantic Web vision is the way Amazon.com Inc. has created an XML version of its database. For some time, Amazon has provided rudimentary tools to equip another Web site to create HTML pages listing books in the Amazon inventory and creating a purchase list that carries back to Amazon's site. The company now provides RDF-like tools for another company's developers to integrate Amazon purchases with their own. Thus, for example, one company could create a single shopping cart with items ordered from Amazon as well as from its own catalogue. Under the term "Web services" many companies provide meta-development tools of the sort that make Amazon's toolkit possible.

The most widely known example of Web service is the Google APIs (Application Program Interface). The Google Web APIs are a free beta service and are available for non-commercial use only, which lets remote applications send search requests to the Google search engine packaged as a SOAP (Simple Object Access Protocol) call. Google executes the request and returns a structured XML document containing the results to the calling program.

6. Evolution of Scientific Communication

One of today's most rapidly evolving and pervasive aspects and challenge for library and information managers is the growth of published material appearing first/only in electronic format. This is particularly true for scientific research publications. More and more authors are placing their papers directly on the Web, traditional publishers are moving into electronic publishing business. The emergence of these novel electronic resources and models has altered the traditional longstanding links between information professionals and their clients. In the past the library was the interface between the user and a big amount of published and unpublished information, whish was made available in hard copy, via online databases or in CD-ROM format.

Nowadays information managers are handling an increasing proportion of new electronic/digital materials. These are produced by the e-publishing industry and, increasingly, by authors using the alternative publishing models in the case of Special and Academic Libraries. This environment has encouraged the emergence of novel publishing models for formal and informal communication among scientists, based on Internet technologies for the dissemination and communication of research materials, with functionalities that far exceed those existing in traditional printing world. First of all, that has promoted a rapid access to scientific information existing in documents, in many cases without a fee. Secondly that also facilitates access to large amount of multimedia materials on the Web and stored in databases, like biological sequences, time series, videos, etc. Some of new electronic publishing models based on open/self -archiving (e.g. deposit of a digital document in a publicly accessible website) – have been tested by scholars in several disciplines and are sponsored by academic departments or research institutions in response to the rising costs in the traditional publishing industry. There is a strong international movement, called Open Archive Initiative that, at least in some scientific areas, seeks to make research papers available by this novel method. Academics and researchers worldwide visit OAI site as the first place to look before deciding to obtain the original document. This is especially useful for countries and/or organizations where financial restrictions cause limited access to wide range of commercially published journals.

7. Benefits Provided by the Publication in *E-Prints* Servers and Misconceptions

The term e-print encapsulates a wide range of meanings. Originally it was defined as electronic preprint circulated among colleagues and field specialists to obtain feedback; the concept of e-print was then generalized to include any electronic version of academic research manuscripts circulated by the author outside of the traditional scientific publishing environment. Journal articles, conference papers, book chapters or any other form of research output can be considered as e-print.
An “e-print archive” is simply an online repository of these materials, which is publicly accessible. Some of e-prints may be peer reviewed before being posted on the servers; others are posted without peer review and authors expect feedback on the results submitted. Typically, an e-print archive is made freely available on the web with the aim of ensuring the widest possible dissemination of its contents to inform colleagues about research in progress and to seek expert comment.
The first e-print server was the Los Alamos Physics Archive, presently known as *arXiv*, it was created in 1991 at the Los Alamos National Laboratory to give access to preprints in the domain of high-energy physics.
In the domain of management, business and finance similar activities have been started creating *RePEC*- Research Papers in Economics service.
Several benefits to be gained from archiving the scientific work of scholars and academics in e-prints repositories have to be highlighted:

- Lowering impact barriers and increasing visibility – papers become freely available for others to consult and cite;
- Rapid dissemination of information to a wider audience;
- Better quality and improved efficiency in the R&D activity (by avoiding duplication) and faster communication between academia and industry
- Improved archiving of scientific data, advantages of the multimedia and the supporting files.

The e-prints offer substantially more features than their print equivalents – annotation facilities, commentaries by pears to be posted. These are benefits from the point of view of the researcher as contributor/reader of the literature.
From the institutional point of view they benefit also by ensuring that their research output is widely disseminated; it helps to enhance their reputation, to attract high quality researchers and to obtain funds for further research.
Furthermore, the e-prints repositories bring added benefits for scientists in poorly resourced organizations or countries. By accessing e-prints repositories available anywhere in the world, they are provided by access to the global knowledge base.
Equally important are the opportunities created by e-prints servers which offer the possibility for scientists in less resourced countries or organizations to distribute local research in a highly visible way and without the difficulties and bias associated with publishing in traditional journals, which tend to favor the publication of papers from well known authors/organizations in more developed countries.

Although besides many benefits mentioned above, some *misconceptions* are still hindering support for e-print repositories. These may be grouped around the following issues:

- Fears that e-prints may give rise to some sort of vanity publishing and consequently have an adverse effect on research quality;
- Intellectual Property Rights (IPR), particularly copyright.

There was a concern that low quality materials will appear on the e-print archives giving rise to some "vanity press" that has not undergone the normal control procedure. However the experience of more than ten years with the *arXiv* repository shows that researchers are always concerned with their reputation and professional credibility and that quality of research has not been at stake.
Furthermore, some e-prints servers have implemented facilities that enable the user to exercise clear options for retrieving material selectively (*OpCit*) to discuss and rank the articles, access the most recent, the most viewed, the most discussed and the highest ranked articles.
The other issue concerning the IPR is rather complex. There is uncertainty as to the ownership of research copyright and the debate is still ongoing. In most higher education institutions accepted custom and practice is that academic authors are permitted to claim and dispose of the copyright themselves. The problem is that commercial publishers of many research journals require the authors to assign copyright to the publisher before publication. A movement taking shape within the e-print community is that authors should be encouraged to retain their IPRs by submitting to journals that do not require signing over the copyright or will agree to the author distributing the papers through e-prints repositories (e-distribution rights).

8. Roles for Librarians Regarding Self-Publishing by Researchers and Electronic Theses and Dissertations (ETD)

The library and information services at universities or research organizations should be *a natural place for e-prints services*. The creation of e-prints archives is a response to a number of structural problems in the academic publishing industry. However, this development does not take place in isolation, librarians must be involved. Strong alliances should be forged between librarians, scholars, scientists and researchers and those that have the responsibility for the development of the infrastructure (Correia, 2002). In this alliance activities for librarians and information managers should be:

- Supporting users (scholars, scientists and researchers) to e-publish their materials, to facilitate self publishing and to smooth the path for potential contributors;
- Exploring the document types that e-prints services can accept and that may be relevant for the nature of research in their organizations;
- Providing advice to those who wish to post their documents on the e-prints services with regard to the evolution of prior publication policies of journal publishers;
- Increasing awareness of the possibilities and facilities provided by e-print archives;
- Persuading institutional managers and policy makers of the benefits to be gained by creation of e-prints services in their organizations.

These are major challenges, according to Correia and Teixeira (2002), representing the future work for librarians and information managers.

Theses and dissertations assume a central role in scholarly communication, as they are sometimes the only tangible deliverables after long and expensive periods of research. As such, they are a major source of new knowledge and contain valuable research results, which are extremely useful to other groups working in the same field.
The creation of digital libraries of theses and dissertations generates an environment, which significantly increases the availability of students' research for scholars and empowers universities to unlock their information resources. This environment gives rise to a number of beneficial activities:

- Improving graduate education – where universities require Electronic Theses and Dissertations (ETD) for graduation they inspire faculty and students to experiment with new mentoring models;
- Empowering students to convey a richer message through the use of multimedia and hypermedia tools, animation and interactive features;
- Training future graduates in the emerging forms of digital publishing and information access;
- Lowering the costs of submitting and handling theses and dissertations
- Increasing accessibility, visibility and readership of students' work;
- Helping universities to build their information infrastructure and extend advanced digital library impact.

The project *Networked Digital Library of Theses and Dissertations* – NDLTD, aims to develop a federation of digital libraries, providing free access to graduate students' theses and dissertations and is a collaborative effort of universities around the world, which promotes creating, archiving, distributing and accessing ETDs.
There are no costs for institutions interested in joining the NDLTD. It is sufficient to send a letter to the NDLTD indicating that intention. Joining only requires agreement with the goals and objectives of the NDLTD. It is essential also to follow the standards developed by NDLTD and make their content available through the NDLTD library (see references).

There are also services providing access on a commercial bases to Theses and Dissertations:

- UMI's Dissertation Services (http://www.umi.com/hp/Support/Dexplorer/
- Dissertation.com (http://dissertation.com)
- Diplomica.com (http://www.diplomica.com)

Academic librarians and information managers have very important future role to perform within their organizations regarding ETDs and other intellectual material produced by academics and scholars, be active promoters of digital scholarship in the Higher Education institutions. Librarians should also be involved into promoting University collaboration at state, national and international level, contributing to the definition and establishment of open, internationally accepted standards that would increase access to the wealth of scholarly information existing in ETDs. They must be aware of these developments in Electronic Scholarly Publishing and prepare their user community to take full advantage of the benefits to be gained.

9. Conclusions

There unmistakable signs of a (r)evolution in scholarly communication – the scientists, researchers, academics and librarians are taking charge to produce information products and services over which they retain control, maintaining their independence from commercial, for-profit publishers. This implies that the information professionals has to become skillful in the use of new applications, including:

- Creation and management of collections integrating resources in a variety of formats;
- Establishing links between library catalogues and the new electronic scholarly materials available on the Internet;
- Address new issues created by the long-term preservation of these new scholarly materials;
- Increasing scientists' awareness of these new sources;
- Supporting potential authors by providing training on electronic publishing;
- Enhance user-created metadata.

Librarians and information managers should be looking for the opportunities, created by the availability of these novel scholarly materials, to ply a significant part in the creation of new knowledge.

References

Aruodai. Resources of Lithuanian cultural heritage – the ethnological, linguistic, folklore, and archeological data base. Available http://www.aruodai.lt/ [6 April 2004]

arXiv – the first e-print service. Available http://www.archiv.org/ [6 April 2004]

Chan, L. and Kirsop, B. (2001) Open archiving opportunities for developing countries: towards equitable distribution of global knowledge. *Adriadne* 30. Available: http://www.adriadne.ac.uk/issue30/oaichan/intro.html [4 June 2002]

CiteBase. Available: http://citebase.eprints.org [6 April 2004]

Cherry, Steven M. Weaving a Web of Ideas. *IEEE Spectrum*. September 2002, 65-69.

Correia A. M.R, J.C. Teixeira. New Initiatives for Electronic Scholarly Publishing: Academic Information Sources on the Internet. Lecture Series on "Electronic Information Management for PfP Nations". Vilnius, Lithuania, 24-26 September 2002

DC-dot – Dublin Core Metadata Editor. Available: http://www.ukoln.ac.uk/metadata/dcdot/ [6 April 2004]

Eprints.org. Available: http://www.eprints.org/ [6 April 2004]

Fensel, D., J. Davis, F. van Harmelen (eds.): Towards the Semantic Web: Ontology-Driven Knowledge Management, Wiley, 2002.

Fensel, D. Ontologies: Silver Bullet for Knowledge Management and Electronic Commerce, Springer-Verlag, 2001.

Google APIs www.google.com/apis/ [6 April 2004]

Nottingham ePrints, About Nottingham ePrints. Value added services Available: http://www-db.library.nottingham.ac.uk/eprints/Nottingham ePrints

NDLTD- Networked Digital Library of Theses and Dissertations http://www.ndltd.org/join/ [6 April 2004]

NDLTD library http://www.theses.org. [6 April 2004]

OAI – Open Archives Initiative. Available: http://www.openarchives.org/, OAI Metadata Harvesting Protocol. Available: http://www.Openarchives.org/OAI/openarchivesprotocol.htm [6 April 2004]

OpCit – the Open Citation Project. Available: http://opcit.eprints.org/ [6 April 2004]

W3C - The Worl Wide Web Consortium. Available http://www.w3.org/ [6 April 2004]

Webopedia – The Online Encyclopedia dedicated to computer technology. Available: http://www.webopedia.com/ [6 April 2004]

Author. Dr. Nerute Kligiene graduated from Vilnius University in 1967 in mathematics, she obtained her doctorate there in 1973. Since 1967 she has been with the Institute of Mathematics and Informatics, where she is currently a senior research assistant. Her scientific interests are statistical analysis of time series and the detection and estimation of changes in model parameters. She is the author of more than 70 scientific articles and the textbook for students. She is an Associate Professor at Vilnius Gediminas Technical University and lecturing in e-publishing at the Vilnius University, she has qualifications of a digital publishing expert. She is one of initiators of the digital publishing training program in Lithuania, the leader of several projects in area of Digital Publishing. In 1996 she joined the UNESCO Chair in Informatics for the Humanities at the Institute of Mathematics and Informatics and became the leader of many Internet projects and interactive CD-ROM production.