New Developments in Electronic Publishing
AMS/SMM Special Session, Houston, May 2004
ECM4 Satellite Conference, Stockholm, June 2004
p. 149-159.

Knowledge Transfer between Mathematics and Industry via the Web: the Math&Industry Initiative

Robert Roggenbuck **Wolfram Sperber**

Zuse Institute Berlin, Dep. Scientific Information Systems
Takustr. 7, 14195 Berlin
e-mail: rroggenb@math.uni-osnabrueck.de *e-mail: sperber@zib.de*

Abstract. Mathematics is a key technology affecting everyday life, e.g., communication, traffic, robots, computers, etc. And many projects in applied mathematics are dealing with the development of innovative technologies and products. But how can a user get a comprehensive overview of and access to the information of a project? Where is the knowledge gained in former projects in this area? There are no services and archives which make accessible all the relevant information in a systematic way. The Web probably gives us an answer to these questions. This paper describes a new attempt to present the information of the projects within the BMBF mathematics program. A complete and deep scheme of the content analysis of the relevant information of a project was developed by using consequently the emerging techniques of the Web such as XML [1], XHTML [2], and RDF [3].

1. Introduction

Mathematics is a key technology for industry and services. More and more mathematical knowledge is inside of innovative products, technologies, and services. Typically, the use of mathematics is a non-trivial challenge for mathematicians and developers in industry and services. Experts from the practice and mathematicians have to work together when they want to solve a real-world problem. At first they must determine a common language to define and describe a problem. Then the real-life problem must be transformed into a mathematical model which is done by developing new mathematical methods and software or by adapting existing ones.

Finally, the results of mathematical modelling must be implemented in the real-life application. Mathematical methods and techniques became more and more important during the last years. Nevertheless the use of mathematics is not visible at glimpse: mathematics is still a hidden key technology.

Mathematical applications are important for

- innovative products, technologies, and services or, in general, for the efficient transfer of knowledge between science on the one hand and industry and services on the other hand
- developers in industry and services who became aware of the potential of mathematics to solve real-world problems
- the public acceptance of the significance of mathematics and hence for the standing of mathematical research and teaching in the future.

Therefore, an adequate and high-quality presentation of the projects is in the natural interest of the mathematical community. Today the Web plays an important role in this respect since it is an essential instrument of presentation. However, the way in which projects of applied mathematics are currently presented on the Web is not sufficient in quantity and quality. Typically, there is no project-related Web site or the information is sparse. Standards concerning the presentation, the structure, the used vocabulary and the technical base for the Web presentation of an applied mathematics projects have been missing so far.

At least two major funding programs in applied mathematics exist in Germany today:

- the Federal Ministry of Education and Research (BMBF) program in applied mathematics
- the German Research Foundation (DFG) Center "Mathematics for key technologies".

Every program covers a lot of single projects in different mathematical fields and application areas.

The BMBF initiated its "mathematics program" in 1993. The fourth funding period of this program covering more than 25 projects now started in 2004. All projects of the BMBF mathematics program are the result of a cooperation between academic institutions and companies in industry and services.

The Math&Industry project [4] that started in 2001 was initiated to develop and implement a concept [5] and method for a comprehensive and efficient Web presentation of the BMBF program in applied mathematics.

2. The concept of the Web presentation

Preliminary remarks

1. Typically, the term Web presentation means a Web site covering one or more html pages. Anybody from all over the world has access to the information contained therein and can get and download it immediately. As a consequence of the great success and rapid expansion of the Web it was necessary to develop methods allowing to handle the information automatically. This requires a special preparation of the Web pages which will be discussed below in detail.

When we speak of a Web presentation this always refers to some aspect of

- the presentation directed to the human user
- a content analysis being the base for the automatic processing of information.

2. Language

   Developers and managers in the German industry and services are the major target group of the Web presentation of the BMBF mathematics program. For that reason, German was chosen as the operating language of the server. Of course this is a limitation for the use of this service outside Germany and it is intended in the near future to allow for it that multilingual presentations are also supported. Since the Web presentations of the projects and the portal are displayed in German we will abstain here from presenting screenshots of the Web pages. Instead the paper focuses on the discussion of the concepts, the principles, the methods, and the implementation of our approach.

The concept of the Web presentation of the BMBF mathematics program bases on a distributed architecture. This means that

- each project builds up its own Web site which is standardized in structure, vocabulary, content analysis, and layout
- central services process the information of the projects and make accessible this information in a user-friendly way.

The following picture should illustrate the relations between the local and the central components:

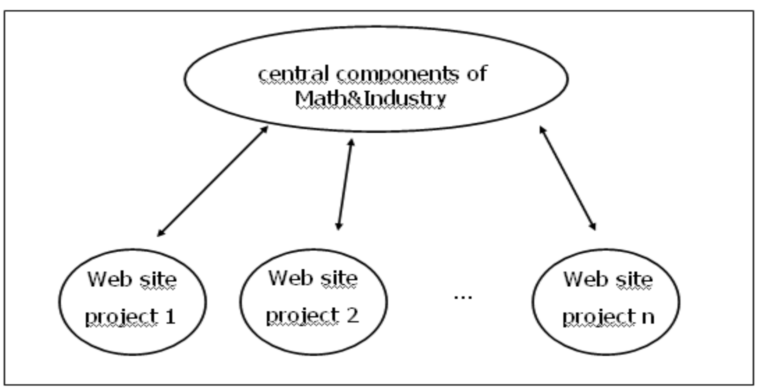

Figure 7: The distributed concept of Math&Industry

The decision pro or con a distributed concept is of vital importance for the organization but also for the technical realization.

Pro's

- The persons working in a project have the greatest knowledge and expertise in the topic. So they are naturally most qualified to implement the Web presentation of their project.
- The projects have full control of the content of their Web presentations.

- The Web sites of the projects can be updated without time-delay.
- Common standards for the Web presentation help reduce the costs of the project Web sites and allow a high-quality machine processing of the information.
- Tools can be developed that make it technically easy and minimize the effort to create the Web page.
- Central services can provide the user efficient access to the overall information.
- Management and maintenance of the central services can be done with a minimum of man power.

Con's

- The projects are responsible for the maintenance of their Web sites on their own which requires particular man power.
- It is necessary to coordinate the activities of the projects and the central services:
  - A personal infrastructure has to be built up.
  - Standards, methods, and tools have to be developed for a distributed concept. The automatic processing of information requires to use emerging techniques.
- The long-term archiving of distributed information has to be organized.

Last but not least, it is the aims and the content of the planned information system that are deciding on the kind of architecture. Our main aim is the complete representation of the relevant information of the projects. It is actually impossible to achieve this goal for projects of the BMBF mathematics program designed according to a central architecture.

In the following the concept is discussed in detail:

- The Web site of a project with its facets
  - A classification scheme for projects in applied mathematics
  - Groups and subgroups: content analysis and metadata
  - Formalization of metadata: the Resource Description Framework (RDF)
  - The WebSiteMaker
- The Portal

3. The Web site of a project

The information on the Web sites of the projects is the backbone of the Web presentation. The Web site of a project should contain comprehensive information about it.

Remark concerning the use of the term "standardization": It was said that a distributed architecture needs standardization to build up powerful services. In the following, standardization denotes a unification of the structure and the semantic description of the given information. It is not a standardization of the content.

3.1. A Classification Scheme for Projects in Applied Mathematics

The relevant information of a project in applied mathematics was analyzed and divided into six groups:

1. Overview:
   This group shall provide a short overview about the project including the general aim, the information for the public. Also a glossary of the project should be part of this group.
2. Applications/products:
   This group shall describe the methods and products, e.g., software which were developed in the project. If results are used in industry and services this should be described here.
3. Participants:
   Here the institutions and companies and the persons involved in the projects shall be listed and described.
4. The problem of the practice:
   This group shall describe the problem of practice and the context of the problem (special requirements, similar problems) in detail.
5. Modelling:
   This group shall present the scientific analysis and the possible mathematical models for the problem in detail.
6. Mathematical treatment:
   This group shall describe the mathematical methods and algorithms that were used to solve the problem in detail. This group should cover also publications about mathematical aspects.

Each group has a number of topic-specific subgroups.

The first three groups provide a quick overview about the projects. Those interested in more details can study the remaining groups. Here one finds detailed information about the problem, its modelling and mathematical approach.

A scheme for the groups and subgroups was developed. It defines the subject of each group and subgroup, fixes the name of each group and subgroup and gives examples for the information in these groups, for more details see [6]. The groups and subgroups define a (hierarchically organized) classification scheme for the information of the project. In other words: the elements of this classification scheme type the information of a project. A classification scheme is a standardized vocabulary for a special topic. So the groups and subgroups define a general metadata scheme for the information of a project.

3.2. Groups and Subgroups: Content Analysis and Metadata

The scheme of groups and subgroups is a first step to standardize the information of a project. But the scheme is not fine enough for our purposes. We need more and specific information about the information of the groups and subgroups to build up the planned central services. For instance, it is important to know that the object is a publication but what is of interest too is the author, the title, keywords, etc. Hence in addition to the groups and subgroups special metadata schemes of the subgroups were defined.

Here, the well-known metadata vocabularies, e.g., the Dublin Core [7] for the description of document-like objects in the Web can be used. However, the standard metadata vocabularies are not sufficient for all groups.

A special feature of the Web presentation of a project is the glossary. This type of glossary should allow an easy familiarization with a project. Moreover, with the help of the glossaries relations between projects can be made transparent (see the overall glossary in the paragraph 'The Portal').

The glossary should list the most important terms used in the problem, the modelling and the mathematical treatment, and provide a definition of the terms and selected interrelations between the terms of the glossary. The interrelations indicate cross-references and different contexts of a term. We have two kinds of interrelations:

- statements on the general context of terms which are well-known from the representation of taxonomies (e.g., the term is similar to another term)
- statements on specific contexts typical for terms in applied mathematics (e.g., a method can be used to solve a problem).

3.3. Formalization of Metadata: The Resource Description Framework (RDF)

As it was said before machines should be able to process the content of the Web presentation of the projects. Therefore the information must be expressed in a suitable and formalized way.

In principle the metadata approach is simple: RDF [3] is – as it is evident from the naming – especially a framework for the content analysis of resources. Resources must have an identifier, e.g., a URL, DOI, or an ISBN, etc. A description of an object can be interpreted as a set of statements of the form "subject – predicate – object", e.g., "A – is the author of – the publication XYZ". These triples define the basic data model of RDF. A triple can be diagrammed pictorially using a directed labelled graph. The arc always starts at the subject and points to the object of the statement. The arcs represent named properties. The graph is a second form to express statements of a resource. Graphs have been used for a long time to make statements, e.g., for the presentation of taxonomies.

Only one question remains open: How can metadata be inserted into the Web presentation? One possible answer is XML: The concept of the Extensible Markup Language (XML) [1] makes it possible to express the triples. Each RDF code can be serialized using the XML syntax.

These are only a few words on the RDF data model and syntax. For more information please visit the official RDF Web pages of the W3C Web site [3].

The following procedure proved successful to formulate the metadata.

1. Define the statements which are important for the automatic processing.
2. Bring the statements in the triple form.
3. Encode the so-defined RDF in XML.

Of course this won't work (the information can't be processed by machines) if there are not used standardized vocabularies for the description of objects. For instance, a machine cannot know whether the predicate 'author' in a statement A means the same as the predicate 'creator' in a statement B.

RDF Schema (RDFS) [9], the Ontology Web Language (OWL) [10], and also the Dublin Core Metadata Scheme (DCMS) [8] define special vocabularies (in technical terms: XML namespaces) for different purposes and applications. If metadata from an existing namespace will be used in RDF code this is expressed by an unambiguous reference to its definition. A machine can then decide whether the relations or nodes in different statements are identical or not. Generally these vocabularies do not suffice to describe all statements that are of interest in real applications. Then the user has to specify his own vocabularies to identify the used nodes and relations in a well-defined way. For instance, Math-Net, see [11], has defined its own RDF vocabulary for the subjects and types of mathematical information. The Math&Industry project applied the Math-Net vocabulary [12] but some extensions were necessary: part of the vocabulary was defined in the metadata schemes of the projects, e.g., to model some special interrelations in the glossary.

In principle, it is possible to integrate the RDF code in an XHTML [2] Web page but up to now this cannot be made in a direct way. Till now, RDF code cannot be embedded into an XHTML page. Therefore the following method is preferred: The Web presentation for the human user should be formulated in XHTML, the expression of HTML in XML. XHTML is identical to HTML expressed in the XML syntax. Each XHTML page is linked with an RDF/XML page containing the metadata.

The presentation of mathematical formulas is a special problem. Typically, mathematicians use TEX as digitizing format. TEX is an excellent print format. But it is not suited for the presentation of mathematics on the Web. TEX must be converted to a format able of being displayed by the browsers, e.g., HTML, PDF, PostScript, etc. However, such conversions may produce detrimental effects, e.g., if TEX is transformed to HTML all mathematical formulas are transformed to pictures. Currently more sophisticated techniques are emerging. Some projects have developed software to transform TEX encoded files into MathML [13] encoded files. These files

- can be embedded in XHTML (XHTML version1.1 and higher)
- can be handled by the major Web browsers.

One such software was developed by HERMES. The result of the TEX conversion is a file encoded in presentation markup of MathML. For more information see [14]. Now, the following scenario could take place if one wants to display mathematics on the Web

- write your mathematical text in TEX
- convert it to MathML
- store it in MathML

The Math&Industry project will test whether this scenario is promising.

3.4. The WebSiteMaker

If you have a look at an RDF XML-encoded document you will find out immediately: this code is unwieldy, long and not intuitive. What you conclude is that it is hard for a normal user who is not familiar with RDF to write correct RDF code and to handle it.

You can escape the dilemma if you develop tools that create the code for you.

Within the framework of the Math&Industry project the WebSiteMaker, see [15] , was developed at [17] and [16]. It is a form-based tool to create a complete Web site of a project or to edit the Web presentation. It can be interpreted as a Content Management System (CMS) for projects in applied mathematics.

The classification scheme is hierarchic and hence the WebSiteMaker is hierarchic as well. After the registration you have to choose the subgroups for which a Web presentation should be created or edited. Then you insert your content or edit it in a form. If you accept the created XHTML page it will be stored.

The WebSiteMaker is installed at a server of [16]. Up to now the created Web sites of the projects are stored on the server of [16] but can be also downloaded to a local server. The central storage of the Web sites makes it easy to archive the Web presentations.

4. The portal

The main aim of the Math&Industry portal, see [18] is to provide a high-quality search and navigation about the information of the projects including different views of special subjects. This aim has to be achieved by using the metadata of the Web presentation of the projects. This means vice versa that the central services can provide good information only if the project Web sites contain appropriate information.

Currently the portal covers different views (lists) of the projects of the BMBF mathematics program. This is a basic service which is expected by each user. Here you can find a survey of the entire projects as well as of the projects carried out in every funding period. A further service is the fulltext search. Therefore all information from the Web sites of the projects is harvested and indexed.

A first special service is the overall glossary. It covers all terms given in the glossaries of the projects. Of course it may happen that a term is defined differently in different contexts. The overall glossary presents

- all definitions of a term ever given in the glossaries of the projects
- links to project glossaries where the term is defined
- the origin of the term (from mathematics, other scientific fields, or practice).

Moreover the interrelated terms and the kind of interrelations are presented. The user himself has to decide, however, whether the meanings of terms are identical or not. Currently the overall glossary is a list that is sorted alphabetically. It is planned to provide further views to the glossary, e.g., a navigation along a classification scheme of the application areas.

An expert database is under construction. This database will list mathematicians that have expertise in special fields of applied mathematics and are willing to support companies with their expertise. The database will list the names of the experts, their field(s) of expertise in applied mathematics, the projects where they are involved, and also their mathematical background. Currently a scheme of application areas is drawn up which is a first attempt to design a standardization scheme of application areas. Another service to be installed in the near future is a product database. The most important product mathematics can offer to industry and services is software able to solve a real-world problem. Often the software is the core of innovative products and technologies. The product database will provide an overview about the products of the projects, the problems that can be solved, the application areas, the developers and contact persons, the existing versions, requirements for use, rights, etc.

Similarly, further services might be provided, e.g., lists of all publications that have been written in the framework of the BMBF mathematics program or a calendar of events. Also maps of clusters of the projects in the different application areas and mathematical fields could be of interest for the user from industry and services.

5. The Web presentation of the BMBF mathematics program – the state of the art

At the beginning there was the concept. Now, the concept is ready and the tools are working. What is required is the information from the projects.

In the mean time, all projects of the BMBF mathematics program have a rudimentary Web presentation. At least the Web presentation consists of the homepage of the Web presentation covering the formal data on the project as the duration period, the participating institutions, the header of the project, the application area and the mathematical field, etc. A lot of the projects of the current (fourth) and the preceding (third) funding period have started to include more information in their Web presentation. Activities (mailing list, workshops) have been started to build up a network of information coordinators of the projects.

The creation of Web presentations of the projects of the first and second funding periods is a particular problem. It makes sense, however, to make this information available too. The methods developed for modelling and solving a real-world problem might be of interest for users in industry and services and mathematicians still today. Unfortunately, it is often the case that the persons that have worked in the former projects have left universities and institutes or work elsewhere. Hence another concept was developed to put the information of the former projects on the Web. The Springer Verlag has published two books about the first two funding periods of the BMBF mathematics program. The Springer Verlag has agreed that the information contained in these books can be used to complete the Web presentation of the projects.

6. Outlook

Building up and maintaining an adequate Web presentation of a special topic means to solve a lot of different problems. The Math&Industry project is a first step as far as projects in the field of applied mathematics are concerned. It will help broaden the information of projects and their results in a significant way.

The concept of the Web presentation that was developed by the Math&Industry project is user-oriented. It is initiated from the needs and the interests of the different target groups. On the other hand the concept makes use of the evolving methods and techniques of the WWW. In its consequence a permanent technical updating and improvement of methods and tools employed will be necessary.

The presentation of the BMBF mathematics program is only a building block for a better Web presentation of mathematics, but it can be the a starting point for more and well-structured information in applied mathematics. Therefore the cooperation with other initiatives in mathematics and also outside mathematics is aspired.

References

[1] Extensible Markup Language XML
`http://www.w3c.org/XML/`

[2] Hypertext Markup Language (XHTML)
`http://www.w3c.org/MarkUp/`

[3] Resource Description Framework (RDF)
`http://www.w3c.org/RDF/`

[4] Math&Industry Project
`http://soft-pc3.zib.de/MathInd/fkz03GRMIB4/`

[5] The concept of the Website of projects
`http://soft-pc3.zib.de/MathInd/HelpFiles/konzept.html`

[6] The schema of the Website of projects
`http://soft-pc3.zib.de/MathInd/HelpFiles/konzept.html\hyper@hashgruppen`

[7] Dublin Core
`http://www.dublincore.org/`

[8] Dublin Core Metadata Schema
`http://www.dublincore.org/dcmi-terms`

[9] RDF Vocabulary Description Language 1.0: RDF Schema
`http://www.w3.org/TR/rdf-schema/`

[10] Ontology Web Language
`http://www.w3.org/TR/owl-ref/`

[11] Math-Net
`http://www.math-net.org`

[12] Math-Net Descriptors
`http://www.iwi-iuk.org/material/RDF/1.1/descriptor/mnd.rdf`

[13] MathML
`http://www.w3.org/Math/`

[14] HERMES
`http://www.aei.mpg.de/hermes/`

[15] WebSiteMaker
`http://soft-pc3.zib.de/create_website`

[16] Zuse Institute Berlin
`http://www.zib.de`

[17] Institut für Wissenschaftliche Information Osnabrück
http://www.iwi-iuk.org/

[18] The portal of Math&Industry
http://www.mathematik-21.de

Received November 25, 2004