Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

WOLFGANG DALITZ WINFRIED NEUN
WOLFRAM SPERBER

Semantic Annotation in Mathematics and Math-Net

online version available under ftp://ftp.zib.de/pub/zib-publications/reports/ZR-03-15.{ps,pdf}

ZIB-Report 03-15 (May 2003)

Semantic Annotation in Mathematics and Math-Net

Wolfgang Dalitz Winfried Neun Wolfram Sperber

ZIB Berlin
Takustr. 7
D-14195 Berlin

17th June 2003

Abstract

The Web of the future will provide a huge amount of information. We need better ways for dealing with and managing the information. A qualified semantic annotation of the information plays a key role for the Web of the future.
This article gives an overview about the efforts of the mathematical community to build up a distributed and open information and communication system for mathematics: the Math-Net. The Math-Net Initiative has developed metadata schemas for some classes of Web resources which are relevant in mathematics. Math-Net Services process this information and enable the user to efficiently search and access the information.

1 Introduction

1.1 Web Resources in Mathematics

The Web is a huge data and information pool also for mathematics. The well-known search engine Google lists more than 10,000 mathematical Web sites. Each of these Web sites provides hundreds of documents as

- mathematical publications (the important mathematical journals, digitized literature, preprints, ...),
- software (numerical software, computer algebra systems, ...),
- teaching materials (textbooks, exercises, examples, ...),
- personal homepages of mathematicians,
- information about mathematical institutions,
- ...

The information is located at different servers of mathematicians, mathematical departments, libraries, publishing houses, software companies, ...
The information is given in different formats (e.g. TEX, ps, pdf, HTML, also Word, special formats for software, ...), and is described by different methods and data sets (metadata).

This is a typical situation in the Web: the information is heterogeneous and located at various servers. And the user is confronted with such problems as

- How can I find and use the relevant information I need?
- What about links to other resources that are important and relevant in this field?
- How can I integrate results from other resources in my work?
- ...

1.2 Semantic annotation

The Web as it is at present cannot solve problems of this kind in a suitable way. The content of the Web documents has to be better structured and marked up ("semantically annotated") to allow a semantic processing of the information. In recent years, the W3C has started various initiatives and activities to develop concepts and methods for the so-called Semantic Web, the Web of the next generation.

One of the basics of the Semantic Web is the term "metadata". From a naive point of view metadata are simple "data about data".
The Dublin Core Metadata set (DC metadata), see [1], is a prominent example for metadata elements which allow a semantic annotation. The DC metadata define a set of 15 elements plus qualifiers for a standardized bibliographic description of electronic documents. The metadata elements cover attributes auch as the title, the creator, or the subject of a resource. In many cases, these metadata are helpful for the content analysis of an Internet resource and for an efficient search.

2 The Math-Net Project

2.1 The aim

The story of Math-Net began with the Math-Net project in Germany (1997-1999). At that time, many mathematical institutions, e.g. departments and institutes, started to build up their own Web sites. The most important aims of the Math-Net project were to

- build up high quality Web sites of mathematical institutions
- make the information searchable and accessible in a user-friendly and effective way.

Especially, the Math-Net project has developed

- methods for the standardization of the semantic annotation for the main types of information offered by a mathematical institution,
- services allowing an effecient retrieval and access to the information.

Building up Math-Net is not only a technical problem, but an organizational, too. A network of persons is necessary to realize such a project. The mathematical departments and institutes in Germany have appointed information coordinators who are responsible for the information provided on the local Web sites.

2.2 Web Sites of Mathematical Institutions and Math-Net Pages

The Math-Net project started with the information provided by the departments and institutes on the local servers of the institutions.
In a first step, the existing Web sites of mathematical institutions were analyzed. The result was not surprising: the Web sites of mathematical institutions have a common core.[1] The evaluation has led to a hierarchical schema for the core information of mathematical departments and institutes. The updated schema of the information objects is the following

- General
  - About us
  - Organization of the Department/Institute
  - Information for Prospective Students
  - Community Outreach
  - Information for Visitors
- People
  - Faculty/Staff
  - Students
  - Long-term Visitors/Associates
- News
  - Schedule of Events
  - Information for Dep. Members/Members of the Institute
  - Positions Available
- Research
  - Research Groups
  - Preprints/Publications
  - Projects

[1] But it was a hard and long-term job to reach an agreement on the standard.

  - Software Development
- Teaching
  - Academic Programs/Curricula
  - Class Schedules
  - Course Information and Materials
- Information Services
  - Computing Services
  - Libraries
  - Journals
  - Bibliographic Search

Core information of mathematical departments and institutes

The groups and subgroups of this list define a classification schema for the information of mathematical departments and institutes. The objects in the classes given by the groups and subgroups have a well-defined subject.
This classification schema was used for the definition of the Math-Net Page. The Math-Net Page is designed for the role of a secondary homepage for the Web site of a mathematical department. The mathematical departments should install a Math-Net Page as a portal to the information offered by them.

2.3 Preprints and Metadata

The medium "preprint" has been a vehicle for the presentation of new results in mathematics for a long time. With the advent of the WWW, preprints became more and more popular in the mathematical community. Publishing preprints in the Web has some advantages:

- Research results can directly be published by the authors in the Web, without time delay and without additional costs.
- Preprints provided on Web servers are accessible from all over the world.

Preprints were the first resource class which was investigated in the Math-Net project.[2] Preprints are offered on the Web sites of mathematical departments, specialized preprints servers organized by the mathematical community, and other institutions. The semantic annotation of the preprints (in form of metadata) differs between various providers.
Within the Math-Net project, a metadata set for preprints was defined basing on the Dublin Core metadata set. The encoding of the metadata was realized in HTML.
The metadata cover

- the author(s) (with the subelements last name, first name, e-mail address)

[2] This is based on investigations made by J. Plümer and R. Schwänzl, see [26].

- the title of the paper
- the URLs of the paper (different formats are allowed)
- the language
- the title of the series, if the paper is published in a series
- the size of the paper
- important dates of the publication (upload, updates)
- classification codes: primary Mathematical Subject Classification (MSC), a secondary MSC, the classification codes used by the Zentralblatt für Didaktik der Mathematik (ZDM) and Computing Reviews (CR), Physics and Astronomy Classification Scheme (PACS)
- keywords
- abstract

2.4 Web Sites of Individuals and Professional Homepages

More and more persons and especially scientists provide information about themselves in the Web. Professional homepages of mathematicians are also an important Web resource. A first model for the description of persons was developed.
Professional homepages of mathematicians are often not only simple Web pages, they are Web sites covering the curriculum vitae, the affiliation, research interests, preprints and publications, teaching activities, ...
Similarly as for the Math-Net Pages of institutions, it makes sense to define standardized professional homepages containing metadata information for persons.
Especially, a professional homepage should contain

- name, affiliation, address, phone, fax, email

and the following groups

- General
- Collaborations and Cooperations
- News and Miscellaneous
- Research
- Teaching
- Professional Societies and Activities

2.5 Tools

Authors need tools to create a Math-Net Page or metadata for preprints and professional homepages in an easy and standardized way. Therefore some software, the Page Maker and the Mathematics Meta Maker (MMM), was developed by the University of Osnabrück and ZIB Berlin.

2.6 Math-Net Services

The institutions taking part in Math-Net make their information resources electronically available in a standardized fashion. They have full responsibility for the quality, accuracy, timeliness, and appropriateness of the data they contribute. Math-Net Service Providers combine these data into services. The Math-Net Services aim at providing fast and well-structured access to the mathematical resources within Math-Net.
Technically, the local information stored on the servers of the Math-Net Members worldwide can be harvested automatically. Machines collect the information from the local Web sites and process it.
A high quality semantic annotation of the objects is the basis for high quality services.
The first Math-Net Service was a preprint index, called **MPRESS** [3] today, the Mathematical PREprint Search System [2].
MPRESS started harvesting of the information about the preprints offered on the Web servers of german math departments. The authors were asked to give a semantic annotation of their preprints described above. Moreover, MPRESS uses some technics for an automatic extraction of metadata.
Here is a list of the current Math-Net Services:

- **SIGMA**[4] is a further Math-Net Service, which is generated by gathering the whole information of Math-Net members, for the URL see [3].
- **PERSONA MATHEMATICA**[5] gathers the information about mathematicians given on the Web sites of mathematical institutions, for the URL see [4].
- **MathJournals**[6] provides an efficient access to the available information of mathematical journals, for the URL see [5].
- The **Math-Net Navigator**[7] collects and processes the information of the Math-Net Pages, for the URL see [6].
- The **Math-Net Links**[8] are a general portal to mathematically relevant Web sites, for the URL see [7].

2.7 The Harvest Software – a technical base for Math-Net Services

The Harvest software, for the URL see [8], is a tool to collect and process data from the Web. After gathering the information will be indexed. In more detail,

- First Harvest collects the data from the servers and works out how to read the data. This depends on the format of data. Harvest supports a lot of different data formats. Summarizers analyze the content, e.g. for HTML, ps, pdf, TEX, plain text, ... Especially, Harvest can be used to collect the metadata information from the local servers.

[3] But MPRESS is no fulltext archive. MPRESS delivers the information about preprints.
[4] is provided by ZIB Berlin
[5] is provided by the University of Cologne
[6] is provided by the University of Osnabrück
[7] is provided ZIB Berlin
[8] is provided ZIB Berlin

- Then the data are converted into a special format: SOIF, the Summary Object Interchange Format. The generated data are the so-called "resource descriptions".
- Eventually, these resource descriptions are stored in a database. This can be done by the Glimpse software (default) or, alternatively, by any other database.

The user interface of Harvest makes the data and especially the metadata of the information searchable. The results provide links to the full-texts.
MPRESS, SIGMA, PERSONA MATHEMATICA and MathJournals use the Harvest software to gather the information of the departments.

3 The Internationalization of Math-Net

Efficient access to the information in mathematics is one of the important issues for the International Mathematical Union (IMU). In 1998, the IMU founded the Committee on Electronic Information and Communication (CEIC) to improve information and communication in mathematics.[9]
CEIC has declared the further development and internationalization of the Math-Net activities a major aim of its work. This affects as well the organizational as the conceptual and technical development of Math-Net.
A worldwide mathematical information and communication system must be distributed and open. So that all mathematicians, mathematical institutions and other information providers have the possibility to supply mathematically relevant information.
The Math-Net Initiative defines the concepts and develops tools for such a system putting on the general developments and trends in the Web. The aims, the principles, and the organization of Math-Net are formulated in the Math-Net Charter, see [9].
Standardized metadata sets are the basis for a distributed information and communication system enabling a cross-linking of resources, and powerful services, and guaranteeing an efficient search.

4 RDF and its use in Math-Net

4.1 RDF

In the Math-Net Project the metadata were encoded in HTML.
To encode the metadata within HTML the ¡META¿-tag in the header of an HTML object can be used. Primarily, metadata are intended to allow a processing of the information by machines. The metadata are positioned in the header and are not visualized by Web browsers.
Semantic metadata sets, e.g., the DC elements, can be expressed in HTML as attribute-value pairs. The DC elements define the attribute. Within an HTML document all values are related to exactly one object, e.g. a preprint.
But the use of HTML as syntax is a strong limitation and only a first step towards a high quality semantic annotation of Web objects.

One example:
A lot of publications have more than one author. Important information about each author is, e.g., her/his name, surname(s), and her/his e-mail address as contact information. HTML has no possibility to group this information. If a publication has more than one author, it can be a problem to find a correspondence between the name, the surname, and the e-mail address.
The Resource Description Framework (RDF), [10], is a general model and syntax for the encoding of metadata, which was developed for a comprehensive and flexible content analysis of Web objects, see [11].
Triples are a convenient method to express semantics. The term "statement" is used in the RDF specification to describe such a triple. Triples are a native extension of the attribute-value pairs which were used in HTML. The subject defines the resource to which the statement is related. The predicate of the statement is the attribute and the object of the statement is the value. It is possible to combine statements.
RDF statements have different representations. A first is the triple representation. A second form is the graph representation which can be visualized if the schema is small. RDF statements can also be expressed in the Extensible Markup Language (XML). The use of XML to express RDF statements combines the advantages of XML for a precise description of the structure of resources with the capabilities of RDF to describe semantics.
The RDF schema, see [12], provides additional possibilities to describe semantic interrelations.

4.2 Use of RDF on the Math-Net Page

Math-Net has used RDF to improve the semantic annotation of mathematical resources. This will now be described in more detail for the Math-Net Page.
The original Math-Net Page developed in the Math-Net project was an HTML document with sparse metadata. These metadata were restricted to

- the title: Local Math-Net Guide
- a subject,
- keywords, e.g., Math-Net, Math-Net Guide, ...,
- and a link to the used Math-Net schema.

Especially, these metadata don't describe the main function of the Math-Net Page: to be a portal to the local resources which are characterized by the corresponding subjects.
In more detail, the DC element relation can be used to describe the linking to the local resources. But it is not possible to add more information to the references, e.g.

- to characterize the subject of the groups and subgroups of the Math-Net Page,
- to assign other labels to groups and subgroups.

Summarizing, the DC/HTML description was not sufficient for the semantic annotation of the Math-Net Page.
For the semantic annotation of the Math-Net Page a customized metadata set was defined. The metadata schema makes statements about

- the groups and subgroups of the Math-Net Page,
- the institution and the information coordinator.

Established metadata sets such as the Dublin Core metadata set or the vCard metadata set for the description of persons and affiliations were used as far as possible.
The semantic annotation of the Math-Net Page was developed in RDF. Therefore a subject and a type schema were introduced:
Each of the groups and subgroups on the Math-Net Page has a well-defined content: this is expressed by the subject schema, see [13].
On the other hand, a group or a subgroup defines a resource class. The resources of each class are characterized by the same subject. The classification is represented in the type schema of the Math-Net Page, see [14]. The subject and the type scheme express different aspects of the groups and subgroups of the Math-Net Page described above.
The so-called descriptors [15] define the groups and subgroups of the Math-Net Page using the RDF Schema Language. The descriptors refer to the subject and type schema described above.
The vocabulary defined by this schema is accessible via a namespace.
The metadata on the Math-Net Pages are expressed in RDF/XML. The consequence: The new Math-Net Page is an XML object. The visible part of the Math-Net Page is encoded in XHTML, the metadata part in RDF/XML.

4.3 Use of RDF on the Math-Net Page - an Example

The use of metadata on the Math-Net Page is illustrated by an example:
The topic “Preprints/Publications” is a subgroup of the group “Research” of the Math-Net Page.
The subgroup “Preprints/Publications” is linked to a resource which covers information about the research papers of an institution, e.g. preprints or articles published in journals.
In a first step, the vocabulary of the Math-Net Page is defined. Therefore so-called ”descriptors” for the groups and subgroups of the Math-Net Page were introduced allowing a smart formal definition.
Especially, the descriptor refer to the subject and type schema which define the subject of the class, e.g., the class “Preprints/Publications”.

The descriptor for Preprints/Publications can be represented in a RDF/XML schema in the following form:

```
<mnst:Descriptor rdf:about="&mnstns;PreprintsPublications">
<rdfs:label>Preprints and Publications</rdfs:label>
<rdf:value>information about research paper</rdf:value>
<rdfs:comment>A class of resources, which contain information
about (a) given research paper(s).</rdfs:comment>
```

```
<rdfs:seeAlso rdf:resource="&iwi;scheme/Subject/"/>
<rdfs:seeAlso rdf:resource="&iwi;scheme/Type/"/>
<rdf:type rdf:resource="&rdfsns;Class"/>
<rdfs:subClassOf rdf:resource="&mnstns;Research"/>
<rdfs:isDefinedBy rdf:resource = "&mnstns;"/>
</mnst:Descriptor>
```

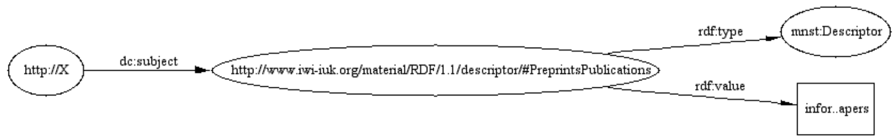
The entries in the type and subject schema are the following

- in the **subject schema**:
  Descriptor: Preprints and Publications
  URI: http://www.iwi-iuk.org/material/RDF/1.1/descriptor/ ↩ #PreprintsPublications
  value: information about research papers
  use with subject: The current resource describes, lists or references resources which contain information about (a) given research paper(s).

- in the **type schema**:
  Descriptor: Preprints and Publications
  URI: http://www.iwi-iuk.org/material/RDF/1.1/descriptor/ ↩ #PreprintsPublications
  value: information about research papers
  use with type: The current resource contains information about (a) given
  research paper(s).

The statement on the Math-Net Page that the resource http://X provides information about the preprints and publications can be expressed in RDF in various representations

a) Graph form [9]



The graph covers statements about resources. Resources have an identifier. They are represented by ovals in the graph.

This graph means (from left to right):
The resource given by the URL http://X has a subject which is given by

[9]The graph was created automatically from the XML representation with the RDF API CARA, see [16].
CARA was developed within the CARMEN project, [23], by the University of Osnabrück.

the resource
http://www.iwi-iuk.org/material/RDF/1.1/descriptor/ ↩ #PreprintsPublications
The type of this resource is given by the resource mnst:Descriptor.
The resource has the value " information about research paper".
The graph uses different vocabularies: mnst:Descriptor is the resource where the Math-Net Descriptors are defined.
The property "subject" is derived from the DC vocabulary (dc:subject).
Type and value are explained in the RDF vocabulary (rdf:type and rdf:value).
The object "information about research paper" is a string (represented by a rectangle in the graph).

b) Triple form [10]

CARA CARMEN RDF API

Parse Results

Subject	Predicate	Object
http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#PreprintsPublications	http://www.w3.org/1999/02/22-rdf-syntax-ns#value	l('information about research papers')
http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#PreprintsPublications	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	r('http://www.iwi-iuk.org/material/RDF/1.1/descriptor#Descriptor')
http://X	http://purl.org/elements/1.1/subject	r('http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#PreprintsPublications')

The triple form of the schema lists the three statements given in the graph representation above.
The subject of a statement must be a resource. The object of a statement can be a resource or a string (literal).
The information if a object is a resource respectively a literal is marked up by the "r" respectively "l".

c) XML representation[11]

```
<rdf:Description rdf:"about=http://X">
  <dc:subject>
   <mnst:Descriptor rdf:value="information about research papers"
     rdf:about="http://www.iwi-iuk.org/material/RDF/1.1/
     descriptor/#Preprints Publications"/>
 </dc:subject>
</rdf:Description>
```

The XML representation begins with the resource which is described.
The second line defines the property.
The next lines contain the object of the first statement, the resource
http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#Preprints/Publications
which has a type (defined in mnst:Descriptor) and a value ("information about research").

[10]The representation as triples was created automatically with the RDF API CARA.
[11]The XML representation was created by R. Schwänzl, University of Osnabrück.

Combining Statements

The semantic annotation of the resource in the example can be extended by the fact that the Math-Net Page provides a link to “Preprints/Publications”. Then the corresponding RDF model is described as

a) Graph form

The statement that the Math-Net Page (the resource which is generated is described by ”online”) references to the resource which is given by http://X is added to the graph.
The predicate “references” is defined in the DC vocabulary.

b) Triple form

CARA CARMEN RDF API

Parse Results

Subject	Predicate	Object
http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#PreprintsPublications	http://www.w3.org/1999/02/22-rdf-syntax-ns#value	l('information about research papers')
http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#PreprintsPublications	http://www.w3.org/1999/02/22-rdf-syntax-ns#type	r('http://www.iwi-iuk.org/material/RDF/1.1/descriptor#Descriptor')
http://X	http://purl.org/elements/1.1/subject	r('http://www.iwi-iuk.org/material/RDF/1.1/descriptor/#PreprintsPublications')
online:	http://purl.org/dc/terms/references	r('http://X')

c) XML representation

```
<rdf:Description rdf:about="http://X">
    <dc:subject>
      <mnst:Descriptor rdf:value="information about
         research papers"
         rdf:about="http://www.iwi-iuk.org/material/RDF/1.1/
       descriptor/#PreprintsPublications"/>
    </dc:subject>
</rdf:Description>
<rdf:Description rdf:about=" ">
    <dcq:references rdf:resource="http://X"/>
</rdf:Description>
```

The examples should illustrate the use of the RDF within the Math-Net Page only. The complete RDF part for the Math-Net Pages is more comprehensive, e.g.,

- It should be possible to assign other labels to the groups and subgroups. This can be modeled by the RDFS element label.

- Besides the metadata of the groups and subgroups the Math-Net Page covers metadata about the institution and the information coordinator, the person responsible for the Math-Net activities of the institution. For this purpose the vCard vocabulary is used.

Of course, within the RDF part the corresponding vocabularies have to be defined. Especially, the Math-Net Page uses metadata elements defined in different metadata sets. The used vocabularies are

- RDF,
- RDFS (RDF Schema),
- DC,
- DCQ (DC Qualified),
- MN (Math-Net Classes),
- MNST (Math-Net descriptors), and
- vCard.

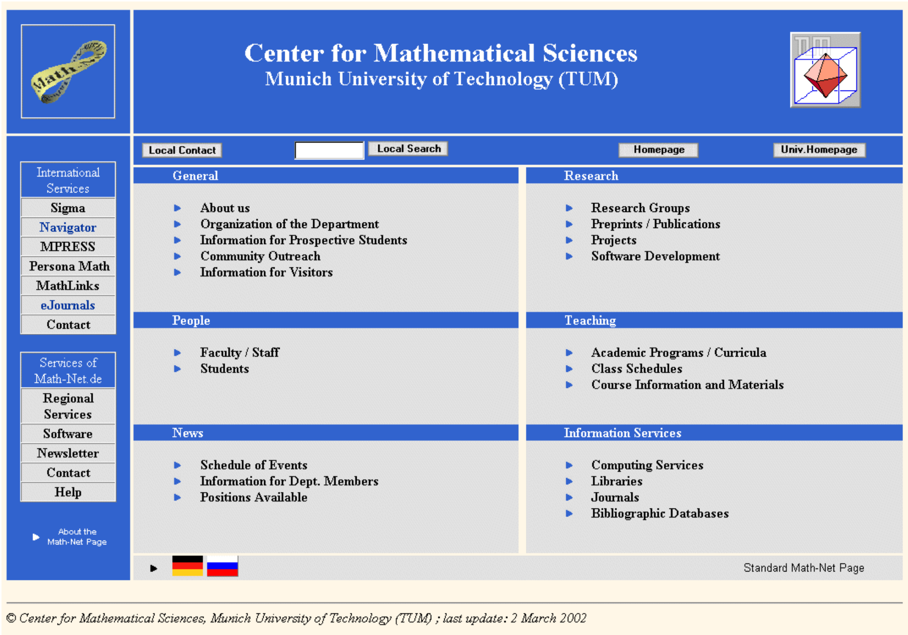
XML uses namespacesto refer to vocabularies. The URIs of the namespaces are termed at the beginning of the RDF part of the Math-Net Page:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
xmlns:dc="http://purl.org/elements/1.1/"
xmlns:dcq="http://purl.org/dc/terms/"
xmlns:mn="http://www.iwi-iuk.org/material/RDF/1.1/Schema/Class/mn#"
xmlns:mnst="http://www.iwi-iuk.org/material/RDF/1.1/descriptor#"
xmlns:vCard="http://www.imc.org/2001/vcard-rdf/3.0#">
```

4.4 More about the Math-Net Page

The following image shows the (Standard) Math-Net Page of the TU München. The left column on the Math-Net Page provides links to the current Math-Net Services, the main part of the Math-Net Page (on the right hand side) covers links to the local information resources of the institution.

The attribute "Standard" means that there is used a controlled English vocabulary. Besides "Standard" Math-Net Pages there are "Alternative" Math-Net Pages with customized labels, possibly in English or another language.
It is expensive to create a Math-Net Page manually. Misspellings in the XHTML part can produce trouble in the layout of the Math-Net Page, errors in the RDF part produce trouble in the automatic processing of the Math-Net Pages.
For that reasons a tool is needed to generate a Math-Net Page in an easy way. The Math-Net Page Maker, see [17], is a form-based tool for this task. The user has to insert several data in a cgi-script via a form.
The script generates the complete Math-Net Page.
On April 12,2002, CEIC and IMU endorsed a recommendation to all mathematical departments and institutes to create and install Math-Net Pages, see [27]. In the mean time more than 180 departments and institutes have generated and installed Math-Net Pages.



Of course, the concept of the Math-Net Page is not restricted to describe a portal to the information of mathematical departments. It can be used to model portals for different purposes. Especially, the idea can be used to develop a proposal for a standardized professional homepage.

4.5 The Math-Net Navigator: Processing of Math-Net Pages

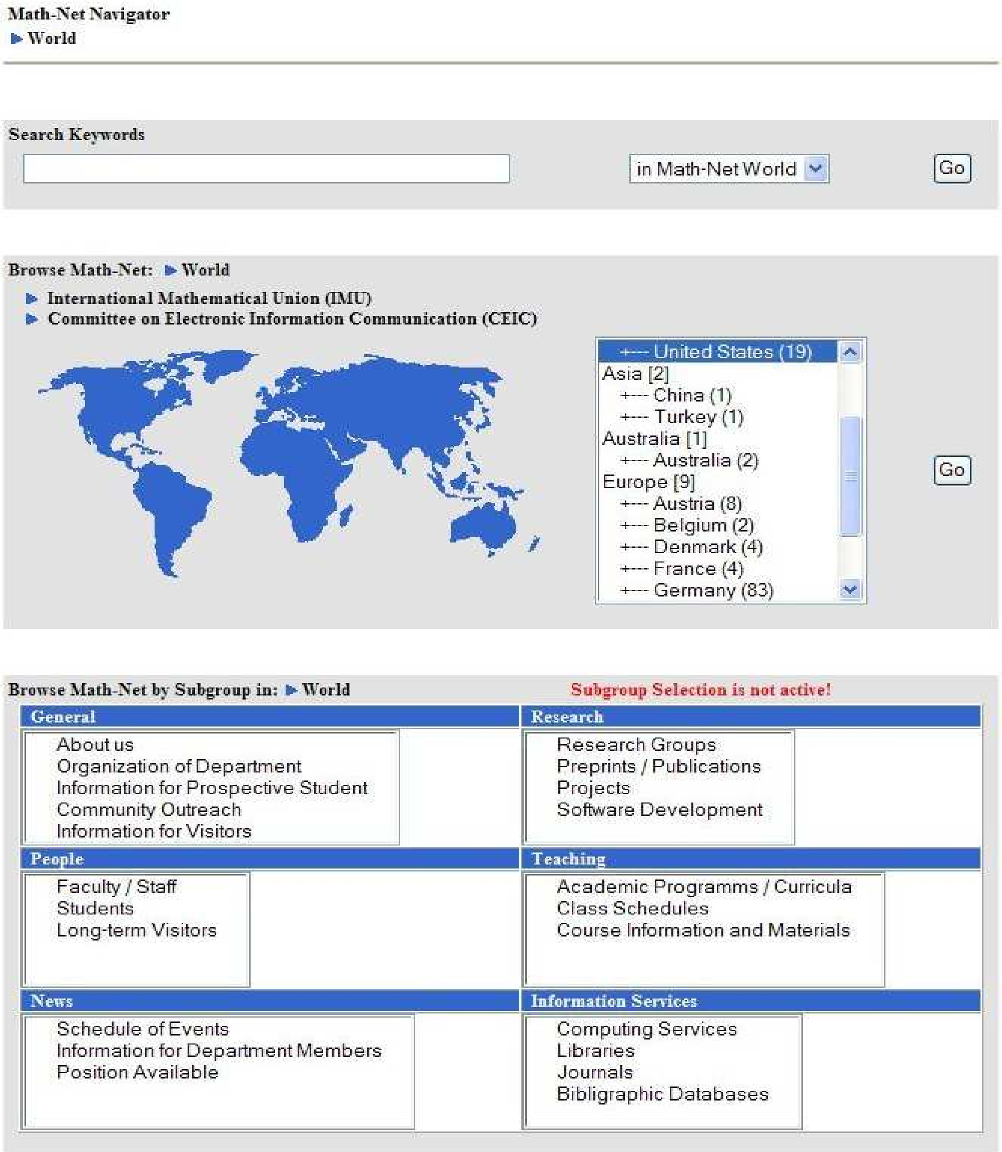
Generally, the RDF metadata in the Math-Net Page allow the processing of the information by arbitrary search engines. Therefore, the search engines must be able to process RDF/XML resources.

The Math-Net Navigator is the Math-Net Service processing the Math-Net Pages. The backbone of the Math-Net Navigator is a database (for that purpose the public domain database MySQL is used).
At first the URL of each new Math-Net Page has to be inserted in the database[12]. The complete Math-Net Pages are daily gathered and processed by an automatic procedure.

The Math-Net Navigator combines seaching and navigation at different levels. In more detail, the Math-Net Navigator provides on the top level

- a search function (over the servers of all Math-Net Members worldwide);
- geographical-oriented views of the Math-Net Members (a clickable map and a list of countries);
- subject-oriented views of the information provided by the Math-Net Members (e.g. lists of all Math-Net Members offering preprints).

[12] Standardized paths for the Math-Net Pages support the gathering of Math-Net Pages. Hence, Math-Net has a URL convention to localize the Math-Net Pages on the server of mathematical departments.



If you choose a country (e.g., Canada) you get, essentially, the same functionality restricted to the Math-Net Members in Canada, namely

- a search function (over all Canadian Math-Net Members);
- a list of Math-Net Members in Canada;
- subject-oriented views of the information provided by the Math-Net Members (e.g., lists of all Math-Net Members offering preprints). [13]

[13] not in the cutout

Math-Net Navigator
▶ World ▶ America ▶ Canada

Search Keywords

in Math-Net World Go

Browse Math-Net: ▶ World ▶ America ▶ Canada

▶ Name	▶ City	▶ State	▶ Type
Departments/Research Institutes			
▶ St. Francis Xavier Univ.	Antigonish		Department
▶ SFU	Burnaby		Department
▶ Univ. of North Carolina at Charlotte	Charlotte		Department
▶ Uni Durham	Durham NC		Department
▶ Uni New Brunswick	Fredericton		Department
▶ Dalhousie University	Halifax		Department
▶ Univ. Lethbridge	Lethbridge		Department
▶ University of WO	London, Ontario		Department
▶ Uni de Montréal	Montréal		Department
▶ McGill Uni	Montréal		Department
▶ Uni Ottawa	Ottawa		Department, List
▶ Univ. of Northern British Columbia	Prince George		Broker
▶ Univ. of British Columbia	Prince George		Department
▶ Mount Allison Uni	Sackville		Department
▶ Brock Univ.	St. Catharines		Department
▶ Memorial Univ.	St. John's		Department
▶ Univ. of New Brunswick, Saint John	St. John's		Department
▶ Lakehead Univ.	Thunder Bay		Department
▶ Uni Victoria	Victoria		Department
▶ Wilfrid Laurier Univ.	Waterloo		Department
▶ Uni Waterloo	Waterloo		Department
▶ Univ. of Windsor	Windsor		Department
▶ Univ. of Manitoba	Winnipeg		Department
Lists			
▶ CAMEL			List

4.5.1 Resume

- The Math-Net Pages and the Math-Net Navigator provide multiple views about the Web sites of the Math-Net Members, especially
  - an effective and user-friendly navigation along the subject and type scheme we defined as "core information" of mathematical departments and institutions, see above.
  - a search about the Web sites of Math-Net Members at different levels
  - a geographical based access to the Web sites of the Math-Net Members

  Universal search engines such as Google, AltaVista, etc., are not able to perform particularly the first view.
- The Math-Net Page is a uniform portal to the local information of mathematical institutions,
- The Math-Net Navigator is a central portal to the Web sites of mathematical institutions.

4.6 RDF Metadata for preprints - MPRESS

The original metadata set for preprints was substituted by an RDF schema in the mean time, see [18].
The RDF schema allows, e.g., a precise description grouping of the authors, particularly of names, surnames, and email adresses as well as of URIs and various formats.
The visualization of the graph metadata presentation is given in [19].
Again, the user needs a simple tool to generate a standardized metadata description of her/his preprint. Version 3.1 of the Mathematics Meta Maker (MMM), see [20], generates this description. The MMM tool produces an XML document containing the metadata which are encoded in RDF/XML.
The MMM tool can be downloaded and installed locally.
MPRESS, [2] , mainly gathers the metadata information for preprints. Today, more than 50,000 mathematical preprints located on more than 150 servers worldwide are indexed in MPRESS. MPRESS provides browsing and search functionalities.
The Mathematical Subject Classification (MSC), see [21], is the structure used for browsing.
Searching combines a fulltext and a metadata search.

4.7 Metadata for persons - PERSONA MATHEMATICA

The concept for the Web presentations of persons is currently discussed by the Math-Net Initiative. The aim is to develop a standardized professional homepage as a portal to the information.
For details of the discussion see [22].
The structure of the professional homepage defines the classes and fixes the metadata. The procedure can be the same as for the Math-Net Page.
The Math-Net Service PERSONA MATHEMATICA, see [4], processes the

metadata to realize efficient search facilities for mathematicians.

4.8 Relations between Math-Net Services

Initially the Math-Net Services were stand-alone services. Of course, the various kinds of documents provided in Math-Net are linked:

- Math-Net Pages refer to preprints and information about persons;
- Professional homepages have links to preprints and the institutions;
- ...

As a consequence, all Math-Net Services should cooperate to allow the user a comprehensive access to the information available in Math-Net. Of course, the various kinds of documents provided in Math-Net are linked together. It starts with simple links:

- Math-Net Navigator and SIGMA
  As said before, the Math-Net Navigator provides also a search about the information of the Math-Net Members. Therefore must be added the URLs of new Math-Net Members to SIGMA to make harvesting these Web sites possible.
- Math-Net Navigator, MPRESS, and PERSONA MATHEMATICA
  If a new Math-Net Page is registered providing links to preprints and to information about the staff respectively this information must be integrated also into the services MPRESS and PERSONA MATHEMATICA respectively.
- Update of the Math-Net Links
  The Math-Net Links should be updated by using the information provided by the other Math-Net Services, particularly Navigator, MPRESS, MathJournals

Moreover, the retrieval of the information contained in Math-Net should allow a refinement in searching, e.g., by types. This can be done, e.g., by a standardized navigation bar on the portal pages. Such a navigation bar can provide links to and/or a searching about the other Math-Net Services.

4.9 The technical base of Math-Net Services: Upgrade of Harvesting/Glimpse-Software

As said before, the resource descriptions in Harvest have a special format: SOIF. SOIF is a flat format. Metadata are presented in the form of attribute-value pairs.

The summarizing, indexing and retrieval components in the Harvest/Glimpse software are not sufficient to process the XML/RDF information and to make the information searchable in an enhanced way. Under the CARMEN project, [23], a prototype for the processing and retrieval of RDF/XML documents was developed by the University Osnabrück and the University Dortmund. This prototype, the X-Harvest/HYREX software, allows processing deeply structured documents, e.g. RDF/XML formatted documents, and combines this with enhanced functionalities of the information retrieval (e.g., stemming, soundex, ...). For more information see [24].

5 Expressing mathematics on the Web

So far, this article dealt with the Web sites of mathematical institutions and a better semantic annotation for some typical and important resource classes. This is a first and important step towards improving the visibility of mathematical information in the Web and developing the basics for a Semantic Web in mathematics.

More in detail, the developed methods and tools addressed the semantic annotation of Web objects on a macro level. But what about the semantic annotation of a mathematical formula in a preprint or a publication?
Therefore it is necessary to analyze the mathematical content of the documents on a micro level. Fortunately, XML can be used for a uniform logical markup of mathematical objects on the micro level too. The advantages: a uniform and extensible platform to express mathematics and mathematically relevant information, to add semantic annotation, and to develop standardized retrieval methods and tools, ...

Up to now,TEX is the most important format to express mathematical terms and statements in digital objects. TEX was the first Markup language. TEX allows an adequate presentation of the mathematical terms and statements. It has well-defined tags for a lot of mathematical functions, but it is not convenient to add further semantic annotation.
Some years ago, the mathematical community under the aegis of the W3C has started to develop the Mathematics Markup Language MathML. The advantage of MathML is that it uses XML as platform. MathML was the first XML specification that was recommended by the W3C. The current version MathML 2.0 is available at [25].
MathML seems to be a concept that facilitates an improved expression of mathematics on the Web. Especially, MathML Content Markup can be used for the automatic annotation and processing of mathematical terms, e.g., by computer algebra systems. Semantic annotation in MathML, however, is not standardized yet.

6 Concluding remarks

Knowledge management in mathematics is at its beginning. The examples described above show that the mathematical community launches a lot of activities to improve the quantity and quality of mathematics information on the Web.
But they also show that it is no trivial task to build up a better information and communication system in mathematics:

- The mathematical community is required to define its specific requirements and to contribute to the mathematical information structure in the future. MathML is an example for the influence of the mathematical community on the future development of the Web.

- The main goal is to improve the access to mathematical information for all interested in mathematical science and results.
- The mathematical community must take an active part in the development of concepts and foundations for better information and communication systems.
- New techniques must be designed in accordance with the general developments of the Internet and the Web.
- A qualified semantic annotation of the Web objects will play a key role for the presentation of scientific information in the Web.

References

[1] http://dublincore.org/documents/dces/

[2] http://mathnet.preprints.org/

[3] http://www.math-net.org/search/germany/

[4] http://www.mi.uni-koeln.de/Math-Net/persona_mathematica/

[5] http://elib.uni-osnabrueck.de/Math-Net/

[6] http://www.math-net.org/navigator/

[7] http://www.math-net.org/links/show?collection=math

[8] http://webharvest.sourceforge.net/ng/

[9] The International Mathematical Union, Committee on Electronic Information and Communication:
Recommendations on Information and Communication, General Assembly, Shanghai,August 2002,
also electronically available at http://www.math-net.org/charter

[10] http://www.w3.org/RDF/

[11] http://www.w3c.org/TR/1999/REC-rdf-syntax-19990222/

[12] http://www.w3c.org/TR/rdf-schema/

[13] http://www.iwi-iuk.org/material/RDF/1.1/scheme/Subject/

[14] http://www.iwi-iuk.org/material/RDF/1.1/scheme/Type/

[15] http://www.iwi-iuk.org/material/RDF/1.1/descriptor/

[16] http://zoe.mathematik.uni-osnabrueck.de/cgi-bin/RDF/parser.cgi

[17] http://www.math-net.org/pagemaker

[18] http://www.iwi-iuk.org/material/RDF/1.1/profile/MNPreprint/

[19] http://www.iwi-iuk.org/material/RDF/1.1/profile/MNPreprint/mnpreprinttest.rdf.png

[20] http://www.mathematik.uni-osnabrueck.de/cgi-bin/MMM3.1.cgi

[21] http://mathnet.preprints.org/msc/

[22] http://triton.mathematik.tu-muenchen.de/~kaplan/talks/Halle_DMV_IuK_2002/ Personendaten in einem verteilten Informationssystem-Dateien/frame.htm

[23] http://www.mathematik.uni-osnabrueck.de/projects/carmen/

[24] http://ls6-www.cs.uni-dortmund.de/ir/projects/carmen/wp7.html

[25] http://www.w3.org/TR/MathML2/

[26] J. Plümer and R. Schwänzl:
Harvesting Mathematics, Euromath Bulletin Volume 2, No.1, 1996,
also electronically available at http://www.emis.de/epub/harvest.ps.gz/

[27] The International Mathematical Union, Committee on Electronic Information and Communication:
Recommendations on Information and Communication, General Assembly, Shanghai,August 2002,
also electronically available at http://www.math-net.org/Math-Net-Recommendation.html