A report on Dublin Core based research information service on mathematics

Kazuhiko Asou

Graduate School of Mathematical Sciences,The University of Tokyo

3-8-1 Komaba,Meguro-ku,Tokyo 153-8914,Japan

asou@ms.u-tokyo.ac.jp

Takako Nakahara

Department of Mathematics, Kyoto University

Kitashirakawa Oiwaketyo, Sakyo-ku, Kyoto 606-8502, Japan

nakahara@kusm.kyoto-u.ac.jp

Takao Namiki

Division of Mathematics, Hokkaido University

Kita 10 Nishi 8, Kita-ku, Sapporo 060-0810, Japan

nami@math.sci.hokudai.ac.jp

Abstract

One of the most important information for mathematician is many informal communications – preprints, seminars and so on.

Recently, such information about mathematical researches on "web space" increases more and more without its typial patterns or central sites. We, mathematician, need the search system and its interfaces designed for our research style. In this paper the authors report the outline of the experimental system called "JMISS"– Japan Mathematical Information Service System – provides such services.

Keywords: *Dublin Core, MathNet.*

1   Introduction

Our motivation are based on the Math-Net project [2], that is a Dublin Core based metadata search system in distributed way. On one hand JMISS intends to join this, but on another hand aims another types – both metadata and full text search.

2   JMISS

We set up one host to experiment the functions of JMISS server, and its location is [1] on Department of Mathematics, Kyoto University.

There are several participants provide research information in seven university. Now we have 102 metadata and 2120 data, they are all HTML based.

2.1   JMISS server

At first, the server gathers the web pages contains only metadata that points the starting URI to get. On this pointed URI, metadata and contents itself may be contained. If there are metadata, URI pointed by them are gatherd as data archive. Secondly, all contents on the URI are gathered and the search index are made.

The server processes the following.

1. The participant register the participant side metadata archive URI through HTTP FORM interface.

2. The server gather all the metadata from the registered URI. To do this "wget" program is used.

3. The server gather all data written in HTML from the URI gathered metadata points to.

4. The server make the search index from metadata and data. From metadata, the Dublin Core field index also be made. We use "Namazu" program to do this because the gatherd data may contain Japanese.

2.2   Metadata based information resources

On the sites or persons provide research information (we call them participants), they should prepare the metadata file points to the URI like this:

```
<meta name="DC.title"
 content="Seminar Info.">
<meta name="DC.identifier"
```

```
       content="URI of Seminar Info.">
<meta name="DC.title"
 content="preprint series.">
<meta name="DC.identifier"
 content="URI of prepeint series.">
```

and register its URI to the JMISS server throughout CGI form.

That is a metadata of metadata. By setting this, participant side can changes its metadata archives location flexibly. After JMISS server gets its file, parse the meta tags and get the URI to mirror.
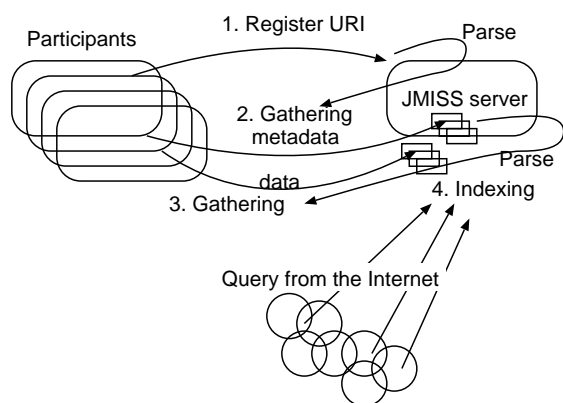

Figure 1. JMISS diagram

3 Examples of metadata based resources

In this section we show examples of participant side resources.

Dublin Core metadata field is made of the following typical informations. So the HTTP FORM interface is useful to make them without human error. Encoding Dublin Core metadata is based on [3].

3.1 Preprint Information

Typical preprint information is like this.

```
HUMATH preprint series

No.2001-1
Title: "Singulaities of Developable Surfaces"
Author: "Author1, Author2"
Keywords: "Geometry"

No.2001-2
Title: "Compact Toeplitz operaors"
Author: "Author Name"
Keywords: "Analysis"
```

Metadata points to a preprint must contain its title, author, series name, series number, format, keywords, and so on. So the typical form is as follows, this form is based on the form of Math-Net:

```
<meta name="DC.title"
 content="A report on using Dublin
   Core for research information
   service on mathematics">
<meta name="DC.Creator.PersonalName"
 content="Takao Namiki">
<meta name="DC.Creator.Email"
 content="e-mail of the author">
<meta name="DC.language"
 content="ENG">
<meta name="DC.identifier"
 content="URI of preprint">
<meta name="DC.subject.MscPrimary"
 content="(Scheme=msc91) 97C80">
<meta name="DC.subject"
 content="Dublin Core, Metadata">
<meta name="DC.subject.topic"
 content="mathematics">
<meta name="DC.type"
 content="preprint">
<meta name="DC.format"
 content="application/postscript">
```

We show only the metadata part and the HTML body may contains abstract and some more informations. So the metadata can be used for keyword search and HTML itself for full text search.

Though the major format of preprint is PostScript or PDF today, MathML, SVG and the other format will be available if they are widely used.

3.2 Seminar Information

Seminar information related to mathematics is as follows.

```
Complex Systems Seminar
Date: 30 June 2001
Time: 16:30
Room: 301
Title: "On Brain Dynamics"
Speaker:  "Speaker name"
```

For such seminar informations, its date,time and seminar room is important. So these details are included and used for field search.

```
<meta name="DC.source"
 content="Seminar Name">
<meta name="DC.date"
 content="2001-6-26T16:30">
<meta name="DC.date.year"
 content="2001">
<meta name="DC.date.month"
 content="6">
<meta name="DC.date.day"
 content="26">
<meta name="DC.date.hour"
```

```
  content="16">
<meta name="DC.date.minute"
  content="30">
<meta name="DC.date.time"
  content="993540600">
<meta name="DC.title"
  content="Seminar Title">
<meta name="DC.type"
  content="seminar">
<meta name="DC.Creator.PersonalName"
  content="Speaker Name">
<meta name="DC.description"
  content="nami.992918121.html">
<meta name="DC.relation.room"
  content="Seminar Room">
<meta name="DC.subject.topic"
  content="mathematics">
```

Abstract and any remarks may be included as HTML body with these metadata. Furthermore, the metadata will points to the "live" data archives in the future.

Though "DC.relation.room" field is just a workaround, not so match the use of Dublin Core, we do not have the alternative way yet. The best method is now under consideration.

3.3 Symposium information

The following is a typical symposium information.

```
Progress of Complex Systems
  August 3-9, 2001.
  RIMS, Kyoto University

Program
  30 July (room 420)
  13:00 - 13:50
    Motomasa Komuro
    GCM and Chaotic Itinerancy
  14:00 - 14:50
    Naoto Kataoka
    Function Dynamics
...
```

We consider a symposium as a set of seminars, the seminar name correspont to symposium title and also the speaker and title.

1. Prepare a detailed data of the symposium.

2. For each talk, make one seminar type metadata points to the symposium data.

This method is easy to imlement, but not so simple. There should be more solution.

3.4 Others

Though there are miscellaneous informations about mathematics, for example, informal lecture notes, mailing lists, web pages and so on, we think we can manage them with metadata as shown above.

4 Conclusion

Now the gathering and search system work well. We think that JMISS as an experiment is successful. Its development itself is further work. The next step is to get more participants and developes the cooperative servers.

One of the problem is to find the method to survey whether the registered contents are suitable for our purpose.

Acknowledgement

References

[1] JMISS. http://diana.kusm.kyoto-u.ac.jp/.

[2] Math-Net. http://www.mathnet.de/.

[3] RFC2731. Encoding dublin core metadata in html.