published in: IEEE Proceedings of ISTAS'02 Social Implications of Information and Communication Technology, pp. 143-149 Raleigh, NC, June 2002

A Research Information Portal for Telecommunications

Kerstin Zimmermann
ftw. (Telecommunications Research Center Vienna), Austria
zimmermann@ftw.at

Abstract

The increasing use of mobile devices and the Internet has led to new research in telecommunications. The combination of various technologies has created interdisciplinary research groups with engineers, computer scientists, mathematicians and physicists working together.

New problems have arisen within these teams. How do we find and exchange the necessary information produced by the entire field working on the same topic but with different views and perspectives? How do we optimize infrastructure and communication as a resource in time and space, and is there one specific way of presenting this piece of knowledge?

1. Introduction

The global scientific community wants to stay in touch with each other through publications or in person. They communicate their projects and latest research results to each other. These contacts expand and transfer expertise .

The exchange of knowledge takes place face to face at work or at conferences where the global research community meets physically and also asynchronously via email. In the latter case mostly data are transferred. The world-wide web becomes the common working environment. Knowledge is still personal, and it is transformed information between human beings [6] as senders and receivers, including faulty transmission.
In this paper we want to present the change of the publication process in the information age, the new defiance of cutting-edge technology as portals and for retrieval from the viewpoint of the field of telecommunication. Knowledge and its presentation are the key factors in this area. The definitions and the ongoing research on how to classify and how to measure knowledge are also an important topic.

2. Publishing

To publish the latest scientific results quickly is no longer a problem nowadays. You only have to put it on your Web server and it's online! You can also add additional information such as raw data, measurements, and the experimental set-up, all of which was and is not possible in a printed- only version.

All institutions and authors provide information about themselves on the Web and wish to disseminate it. They want to inform, to attract attention, to present and to document their work. This can also be seen as a kind of self-representation characteristic of publication culture in research.

In the paper age academic research journals enhanced the scientific value of a publication through the refereeing process. There was a chain of delays, many individuals were involved such as the editor, the referees, the publisher, etc. and lastly the librarian, who collected and categorized the paper again. During this lengthy process the content was transformed many times, cf. [8]. On the other hand, preprints were mailed to colleagues to inform them in advance and to shorten the time before the information was actually published as a reputable paper.

In the E-age, publication (distribution to an undefined group of persons) can be carried out instantaneously via the Web, either by using the author's or his/her institution's Web server or by sending it to the ever-growing number of E-print archives such as ArXiv, started in 1991 by Paul Ginsparg for physics and related fields, MPRESS of Math-Net, begun by M. Groetschel et al., and Citeseer by S. Lawrence et al.
The growing body of scientific information is only a small portion of the data circulating on the Internet. Here a relatively homogeneous group supplies various kinds of information online.

But E-publishing has some additional requirements. An undocumented electronic file may get lost on the Internet. Nowadays it should include metadata such as that about the author, the title, date, format, keywords, etc. In printed publications preprints are distributed, the author is not

concerned with the format and can look up keywords and categorizations in the library if necessary.
The separation of syntax and semantics get requisite. Is necessary for additional changes, and structure elements such as descriptions of figures such as tables, graphics, etc. are essential for full text research. A copyright statement should also be obligatory.

3. Retrieval

Publishing your own information is one thing. Searching for information in the relevant fields of other working groups is another. If we seek to obtain new information about an interesting topic or one with which we are not familiar or just want to update our knowledge of the state-of-the-art, we would not want to use global search engines such as Altavista, Google or Lycos with many irrelevant hits. If, for example, we only know the author, we may want to search for publications or administrative data. If we know the research group, we may want to have a look at the institution directly.

Why is the retrieval of desired information so haphazard? Let's look at the history of the Worldwide Web. In the beginning (1990s) the Internet was the connection between special computers to transfer scientific data. Only text mode with a command line (blue screen) was available, and one had to be a programmer to use the system. Then the first applications such as the first Mosaic Browser and graphical interfaces appeared. The Worldwide Web attracted more and more users, and within a short time moving from research only to include business, too. Other contents were put online, too, including all kinds of illegal data. Multimedia elements provide greater opportunities (including games) and other software such as Flash and Shockwave became common. Lastly other applications such as E-commerce play an important role.

First the area grew without restrictions. But now legal aspects are being considered, an Internet etiquette is recommended, and various standards have been established such as ISO 9241 for user-friendly interfaces.

Information overkill is still there so a defined search space is needed to be certain that the relevant sites are found and indexed (precision and recall).

4. Requirements for Portals

We would like to have one starting point for one (natural) field. This also helps us to stay in touch with the worldwide community and delivers structured information from around the world suited to our individual needs. The information should, of course, be reliable, authorized up-to-date and complete, too. The source should enable us to contact the person associated with a topic in which we are also interested, allowing us to gain administrative and personal information (who works where, telephone numbers). All these information needs require new information management.

A structured information platform is required where the different players can retrieve the data in a particular way. Each discipline uses its own classification scheme such as AMC in computer science, MSC in mathematics and PACS in physics. Generally the author knows only one scheme. In telecommunications research, descriptions are usually very formal so publications consist of plain text, formulas and figures. In engineering or physics there may be additional experimental data and construction plans or software algorithms. Research on similar topics often exhibits one and the same scheme.

For interdisciplinary searching all nominations/ search items may be useful. Homogeneity is not easy to achieve. Correlations within the classification schemes are necessary. In the near future more widespread use of metadata will greatly improve the retrieval results for professional use in research.

Planning a non-commercial portal can be based on the experience of Math-Net and PhysNet, two worldwide information portals under the auspices of the respective research societies.

5. Existing Portals

There are already numerous information portals which address specific interests, but in general these are designed for commercial rather than academic needs.
In this paper we do not refer to non-scientific servers. Only some commercial servers with their services for physics are mentioned:
Sirius[1], owned by Elsevier, has a good intuitive way of providing search results, but the more search items are listed, the less the user gets. PhysLink.com[2] is overloaded with banner commercials including some very adult content so that it is not suitable for concentrated research work. Recently VCH-Wiley/Germany as a late addition to the scene, has tried to set up a physics 'portal' with advertising called pro-physik.de[3].

[1] http://www.scirus.com
[2] http://physlink.com
[3] http://www.pro-physik.de

published in: IEEE Proceedings of ISTAS'02 Social Implications of Information and Communication Technology,
pp. 143-149 Raleigh, NC, June 2002

Commercial science publishers have free entry pages for physics: Examples are Springer Physics & Astronomy[4]. Finally, professional national physics societies have physics entry points, which are free of commercials and are professionally knowledgeable, most prominently the American Physical Society (APS)[5], the European Physical Society (EPS)[6] and the Institute of Physics (IoP)[7] .

Related to telecommunications there is IEEE Xplore[8] by the Institute of Electrical and Electronics Engineers, Inc. and IEE[9] of the Institution of Electrical Engineers.
Accessing the full texts of these publishers is liable to charges, like the INSPEC database for physics, electronics and computing or the Science Citation Index (SCI)[10].

Free online archives as mentioned above are:

- arXiv.org[11] in the fields of Physics, Mathematics, Nonlinear Science and Computer Science,
- MPRESS[12] for Mathematicians and Computer Sciences
- CiteSeer[13] for Computer Sciences

In the online archives you find preprints and documents. The publisher's information portals offer news and edited papers but at both have no additional information about the author or research group. Therefore other types are required which are named below.

5.1. PhysNet[14]

Let us look at the front end of an existing international information portal, PhysNet (figure 1).

Figure 1: Home page of PhysNet (screenshot)

It offers administrative as well as personal data in *PhysDep*, such as the addresses of all physics institutes around the world together with its scientific staff. Individual publications and documents are listed ordered by research groups in *PhysDoc*; published papers are thematically sorted in *Journals*. Collections of services are referred to in *Conferences* and *PhysJobs*. Teaching material is listed under *Education*. Different other tools are listed under *Services*. The last category *Uploadform*, offers interaction to add to the system what is missing. People are free to browse through every category to retrieve its data or search across the whole system.

PhysDep contains 2330 links round the world, 173 in Germany and 370 USA. The collected information is from about 1,500 document lists of local physics institutions and departments: 430 in Germany, 80 in France, 170 in UK, and 278 in the USA and 542 in other countries worldwide. The number of the linked publication lists in PhysDoc seems to be stable with about 1,100 links, but certainly is still incomplete.

Now we change from the front end and user site to the back end to give additional information:

The estimate number of reached documents thus far is well above 80,000 (70,000 in February 2001, 62,000 in April 1999 and 50,000 in October 1998). To keep the links for PhysDoc up to date is time consuming, because links are often moved around without notice by the authors at their local sites. Currently there are 800 requests per day. The service was set up in 1995.

In its early development, PhysNet was an information service from Physicists to Physicists. But later on, the 'provider' of PhysNet wider its audience to students, industry and public who are using science information

[4] http://www.springer.de/phys/index.html
[5] http://www.aps.org
[6] http://www.eps.org/
[7] http://www.ioppublishing.com
[8] http://ieeexplore.ieee.org/
[9] http://www.iee.org/
[10] http://www.isinet.com/isi/products/citation/sci /
[11] http://www.arxiv.org
[12] http://mathnet.preprints.org
[13] http://citeseer.nj.nec.com/cs
[14] http://physnet.uni-oldenburg.de/PhysNet/physnet.html

systems only occasionally. In Germany this was a response to the decreasing number of students (referring to the bad / unattractive image of the natural science among public). Thus good promotion and public relation are needed. Once more, physicists, employed in public research institutions, knowing the techniques of programming, due to their original profession, developed new non-commercial services, taking into account the specific needs of their scientific community, and their low budgets. They also set up some kind of standard, reflecting the international development in W3C and the DC initiative as well as taking part in the discussion of the working group DC.Rights. Keeping in mind their specific needs and requirements they fulfil the demands of their community and build virtual teams.

Since 2002 it is OAI[15] compliant and there are ongoing research project with the Digital Library of Virgina Tech see [10].

The didactics of the learned fields, teaching material, etc. was neglected for a long time. So one could find only very few proven and useful tele-teaching/ long distance learning materials online until now. This field is now rapidly growing.

The open link lists of PhysNet ensure a transparent structure. By seeing the URL you can transverse or bookmark it for the next time, going directly to the piece of information you are interested in. It shows the person who is responsible and gives the contact information because no central database has been set up. Another aim is the flat hierarchical structure but offering one good starting point for information. A site map attempts to present the content. As you can imagine an experienced user may not ever use the developed structure but navigate on his or her own.

PhysNet is demonstrating that the web allows distributed heterogeneous databases, that conjoin as one homogenous service but integrate all relevant information. It is under the auspices of the European Physical Society (EPS) and therefore certified.

A charter for PhysNet has been published which is to be signed by those institutions joining the organization of the content.

5.2. Math-Net[16]

Math-Net as an information portal for mathematics offers various services, such as searching for (*Sigma*) and in (*Preprint*, MPRESS) documents, personal and administrative data in *People* and *Institutes*, *Software* and *Links*. *Info* give background information about the project itself and the *Contact* button allows some feedback, as shown in figure 2.

Math-Net is the analogous service for mathematics, originally based on an eprint- server (see above). It became a portal in several projects after 1997. It now has about 800 requests per day (nearly the same as PhysNet).

What both services have in common is the underlying structure: decentralized servers that build a network; and bare indexed by a shared software tool called HARVEST.

Harvest consists of two parts: the gatherer and the broker. The gatherer collects and extracts all data in registered servers and directories. This is the main difference to global search engines, which act like spiders from a starting point. The broker is the query manager which indexes the data and builds the interface for the search. One broker can request more than one gather or other brokers to process the query.

Figure 2: Starting page of Math-Net (screenshot)

5.3. Comparison

If you compare the two entry web pages you notice the differences: There are two expanded services with different interfaces optimized for specific users. At PhysNet there are 9 categories. In Math-Net there are 8. The number is nearly the same but they can not link one to one because of similar but not identical content.

15 http://www.openarchives.org/

16 http://www.mathnet.de/

The Learned Societies uses their own classification leading to other retrieval methods.
All known physics server are linked in PhysNet. To optimise the task institutions are advised to install their own HARVEST-gatherer and list up their relevant file structure there because you can not find the same information like staff or publication lists at the same place on each server. To prevent a complete breakdown of the system the structure is mirrored by other Physics institutions at eight other servers around the world which redirect the online request in a case of a crash.

As you can imagine from the structure of portals, these are 'democratic' and participatory. Great care has been taken to assure that each institute and national branch enters on an even level. Each institute and national branch is given the chance to edit and maintain their information by themselves. This is realized with linked-list mirrors, distributed gatherers and importing of linked-lists.
In principle the distributed work task allows infinite work sharing around the world to supply the information and assure the service.

Setting up a non-commercial portal can be based on the experience of Math-Net and PhysNet two worldwide information portals under the auspices of the respective learned societies.

6. Telecommunications[17]

Let us consider the conception of an academic service for telecommunications. In contrast to the two other portals figure 3 does not show the starting page due to unfinished layout. Here 9 categories are chosen as well. Some nearly identical to the categories in PhysNet like research institutes, documents / infos and conferences but there are also some different like standards, legal and provider.

Have a closer look at the research institutes site which is still under construction. At the moment (March 2002) it contained 199 links to research groups around the world sorted alphabetically by country domain and city.

[17] http://userver.ftw.at/~kerstin/telecomportal/

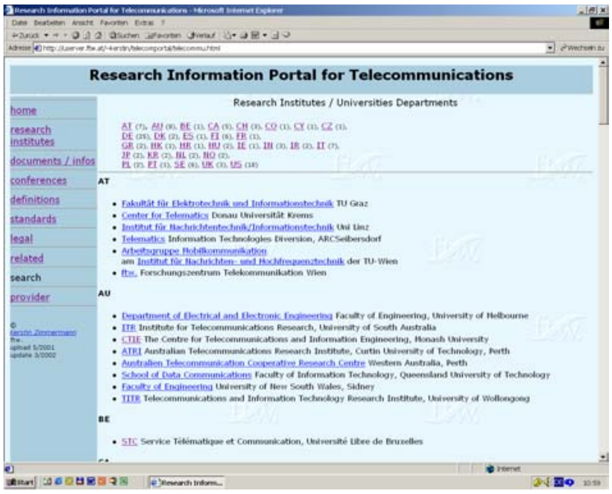

Figure 3: List of the research institutes in telecommunications (screenshot)

What are the difficulties in compiling this lists compared to Mathematics and Physics? The is not a special department / faculty hosting telecommunications but a wide field of research groups located e.g. at:

- Computer Science
- Communications systems
- Communications Engineering
- Electronic / Electrical Engineering
- Information technology
- Mathematics
- Signal Processing
- Telematics
- etc. …

So you have to guess where you may find the information.

In the worst case you have to look at all of them. The same happens if you try to categorize documents or research papers. There are many different classification schemes depending on the learned field like IEEE, IEE used by the engineers. *ACM* is the computer classification scheme of the Association for Computing Machinery, *IPC* the International Patent Classification, *MSC* the Mathematics Subject Classification (2000), *PACS* the Physical and Astronomy Classification Scheme and finally *DDC* the Dewery Decimal Classification used in libraries.
How to manage this? Conceptualising a new one could be an idea but it's unrealistic. It would be an another isolated application not stimulating interdisciplinary research.

Thinking of cross linking and concordance schemata to transfer from one classification to another scheme is another thing but going into detail you'll find:

ACM
B.4 Input/Output and Data Communication
H.3 Information Storage and Retrieval

IPC H 04 Electric Communication Technique

MSC
94Axx Information and communication, circuits
94Cxx Circuits, networks

PACS
01.20.+x Communication forms and techniques (written, oral, electronic, etc.)
84.40.Az Waveguides, transmission lines, striplines
84.40.Ba Antennas: theory, components and accessories
84.40.Ua Telecommunications: signal transmission and processing; communication satellites
84.40.Xb Telemetry: remote control, remote sensing; radar
89.70.+c Information Science

DDC21
003.5 theory of communication and control
003.54 information theory

There are not only different schemes, but also different hierarchies and numbering (letters mixed with numbers, number mixed with letters, only numbers). Setting up a concordance schemata is an enormous task. It can either be done through intellectual input of persons who are familiar with at least 2 schemes or by a software algorithm if there is a test bed with doubly classified documents. Research was done here between MSC and PACS within the CARMEN[18] project: Content Analysis, Retrieval and MetaData: Effective Networking.

7. Information / Knowledge Management

In general, a portal is the output of an information management process. First data must be collected, structured as information and presented. It should also be retrievable and searchable by linking or storing. Human power is necessary to run and feed this service. Planning and scheduling are the first tasks. Efficiency and effectiveness are important as well as maintenance.

Intellect is essential to become an authority in academic life as well in the nomenclature of the web see [2]. A collection like an information portal can then be turned into a test bed for a software algorithm [1] of data mining to detect and collect wider information resources.

The great question of measuring knowledge assets is at the very beginning. For first result and a comparison of different models see [7].

At ftw. there is the challenge of combining commercial and academic research due to the opportunity for K plus centers in Austria.

8. Acknowledgements

This work is supported within the Austrian competence center program Kplus and by the member companies of ftw.
I thank my former colleagues Eberhard R. Hilf, T. Severiens at the Institute for Science Networking (ISN) in Oldenburg, Germany.

9. References

[1] Fürnkranz, J. *Hypertext Classifikation: A Case Study in Hypertext Classification,* Technical Report OEFAI-TR-2001-30

[2] Kleinberg, J., Lawrence, S.: *The Structure of the Web*, Science, Vol. 294, p. 1849, 30 November 2001

[3] Lawrence, S.: *Online or invisible?* Nature Vol. 411, Nr. 6837, p.521, 2001

[4] Lawrence, S.: *Access to scientific literature* The Nature Yearbook of Science and Technology pp. 86-88, 2001

[5] Lawrence, S., Bollacker, K.; Giles, C. Lee: *Indexing and Retrieval of Scientific Literature* , Eighth International Conference on Information and Knowledge Management, CIKM 99, pp. 139-146, 1999

[6] Lugger, K.-M., Kraus, H.: *Mastering the Human Barriers in Knowledge,* Proceedings of I-KNOW'01, J.UCS 1 2001

[7] Bontis, N., *Assessing knowledge assets: a review of the models used to measure intellectual capital* IJMR, Vol. 3, Issue 1, pp.41-69, March 2001

[8] Zhao, J. L., Resh, V. H.: *Internet Publishing and Tranformation of Knowledge Processes,* Communication of the ACM Vol. 44, No. 12, December 2001

[9] Hilf, E., Hohlfeld, M., Severiens, T., Zimmermann, K.: *Distributed Information Services in Physics* High Energy Physics Libraries Webzine, issue 4, June 2001 < http://library.cern.ch/HEPLW/4/papers/2 >

[18] http://www.mathematik.uni-osnabrueck.de/projects/carmen/index.en.shtml

published in: IEEE Proceedings of ISTAS'02 Social Implications of Information and Communication Technology, pp. 143-149 Raleigh, NC, June 2002

[10] Goncalves, M. A., France, R. K., Fox, E. A., Hilf, E. R., Zimmermann, K., Severiens, T.: *MARIAN: Flexible Interoperability for Federated Digital Libraries* (paper submitted to ICDE 2001) <http://www.dlib.vt.edu/projects/MarianJava/icde2001.r8.pdf>