Standardizing Subject Term Assignment for a Library Personalization System

V K J Jeevan   P Padhi

Abstract

The Central Library of the Indian Institute of Technology, Kharagpur is engaged in developing a single window seamless access system for personalizing electronic contents to its faculty and researchers. Since a computer-based retrieval system requires carefully selected search terms to selectively separate useful and relevant information from the large collection of information available in a specific branch of knowledge and most of the electronic systems use technical terms in vogue in the concerned field, sticking to these standard terms while conducting the search helps to avoid the noise and irrelevant retrievals. Since thesaurus and other subject heading lists are very exhaustive and are not very much required in a research area description context, we look for alternative sources, but standard and authentic to create standard subject terms for use with the personalization system. As part of the system, we are devising a subsystem for assigning standardized terms for selecting and incorporating subject specializations of researchers. For example, with the aid of subject terms gathered from the specializations of the 34 Special Interest Groups of ACM and the research areas, courses and projects of faculty members in the relevant departments, we tried to arrive at standardized representation of subject terms for researchers in Computer and allied disciplines and those having computer related specializations in other branches. A similar attempt is being made to extend the subject specializations of researchers in Physics with the exhaustive list of terms provided in the Physics and Astronomy Classification Scheme. This paper will briefly illustrate some personalization projects in USA and UK, contents identified for personalization in the Institute, standardization of Computer Science terms and a partial list for terms generated for Physics along with some of the challenges and benefits.

Keywords : Library Personalisation; Content Personalisation; Search terms; Subject terms

0.   Introduction

The Central Library of the Indian Institute Technology, Kharagpur has been constantly witnessing the widespread availability of electronic information resources on the one hand and the frustration and demand on the part of users to search and retrieve relevant information from a large mass available on the other hand. The Library has initiated several services such as current arrival list of journals, a portal on its website to access the different electronic journals available as per the subscription to the printed counterpart and as part of the MHRD's INDEST consortium, electronic library facility to access electronic and multimedia resources, selective dissemination of information from electronic bibliographic databases to alert and update researchers, etc. These initiatives though helped the user to identify the most useful resources required essentially for their wide ranging academic activities, but never reduced their difficulties to retrieve them as the users always wanted a single window system for seamless access of contents. To achieve this lofty goal and impressed by the success of many personalization projects in libraries and other situations, the library is involved in developing a web based virtual system to personalize the electronic contents of interest to researchers in the Institute. A prototype of the system is already under development and is under testing for trial launch soon.

Any computer-based retrieval system requires carefully selected and workable search terms to selectively separate useful and relevant information from the large collection of information available in a specific branch of knowledge. Most of the electronic systems use technical terms in vogue in the concerned field and it will be always helpful if one sticks to these terms while conducting the search. Also focusing on standard terms helps to avoid the noise and irrelevant retrievals one normally encounters in electronic retrieval systems. The personalization system under development tried to gather the initial input about the subject specializations of researchers from the user's web site, as we are lucky enough to have detailed homepages of faculty members discussing their areas of research, publications, research projects, etc. However, the subject specializations mentioned are to be polished and in some cases need to be laced with context and qualifiers and related and linked with other similar or broader terms to improve their utility while designing a computer based retrieval system. Since thesaurus and other subject heading lists are very exhaustive and are not very much required in a research area description context, we look for alternative sources, but standard and authentic to create standard subject terms for use with the personalization system.

1.   Library Personalization System

Information and its massive growth ahead of the systems to track their access alienated many users from useful contents leading to speculations about under use of information. The spread and inclusion of web on all areas of activity prompted the clients to make wishful comparisons about web and libraries. Users are more aware of the availability and benefits of electronic information and what they expect and even demand from libraries is a web like search and access at the click of a link facility for information resources stored in the library also. Personalizing libraries is projected as a viable alternative to tide over the information explosion, and to conserve the time of users in more productive intellectual tasks, in the aftermath of large quantum of e-information and the emergence of affordable cool web technologies to efficiently manipulate them as per the individual needs of users. Personalization is a mechanism to tailor recommendations in real-time to application users, during a user's application session [2]. Providing special treatment in the form of information and applications matched to a library user's interests, roles, and needs is known as personalization [3]. Choosing the most relevant content to deliver, based on past user behavior as collected in the user profile and embedding this personalized content into application output or Web pages in a flexible manner are regarded as two key functions of personalization [2].

Electronic information and its easy search and access features facilitated many service providers to profile and select contents based on various attributes of the service seekers. The technological advancements made by web enabled computing and the continued interest of users to gather information resources on the web enabled many of the popular portals to bring in the concepts of personalizing their contents to suit the varied interests of users. It is not only libraries, but there are several other service providers who resort to the concept of personalization to bring in more users and site visitors. The electronic commerce and related applications also give more emphasis to identify and recognize individual users so that information regarding services of interest will be provided to potential clients at frequent intervals leading to more volume of business and more satisfied customers. Apart from selecting contents, some services also provide options to select the appearance of the pages like background colours, fonts, order of menus, etc.

1.1   Illustrative Examples

Rediff.com provides personalization services, covering news personalization, matchmaking, astrology, shopping, etc [5]. A user can get personalized My Yahoo! page by choosing exactly what is needed from different information tools like news, email, weather, entertainment, shopping and stock quotes and where to put it by organizing them in the way the user likes [6]. By creating a login in, a user gets a specialized view of recommendations, new releases, best bargains, etc., from the famous online shopping portal for books and several other items, Amazon [7].

In the context of libraries, through personalization, users start with a generic library and by making selections and choices create a personal library with pre-selected and relevant e-journals, databases, reference works, websites and other information resources. Also the user can view the different services offered by the library through a single window implementation strategy (there may be different systems for in-house resources and network accessed information, however they appear as a single system to the user) with provision for configuring alerts and automatic updates and users can add/remove resources of their choosing. MyLibrary @ Notre Dame University, USA attempts to resolve the issues raised by users that information is hard to find; information is hard to get; and for improved communication between library and users [8]. With a relational database of links to full-text databases, data sets, bibliographic databases, and electronic texts of all kinds, a Web server, and a set of scripts, Mylibrary @ North Carolina State University, USA, provides a user-driven, customizable information service that allows user to create a portable Web page listing information resources available by telling a bit about him/her [9]. The BUILDER at University of Birmingham, UK aims at 'seamlessness' access of all resources through the same interface and through profiling resources of relevance are offered as per the needs of the individual user [10]. MyOpenLibrary Project at Open University, UK concentrates on seamlessly integrated, user-focused, interface for users to access their personal selection of Library resources with linkage to other departmental and University-wide information systems [11].

2.   Contents for Personalization

Some of the services offered as part of personalization such as, details about books issued, renewed, and reserved by users through the Library Automation system, electronic notice boards to discuss events and academic work in the institute, etc., need not require subject specialization from the users and these can be offered without any direct input, whereas services such as OPAC of library holdings, email contacts to librarians and other staff, recommendations for book and journal purchases, requests for ILL/DDS services, etc., require variable input from users depending on the information need at a particular instant. Services rendered in a semi-manual or manual approach are to be automated in the first place to interface them with the personalization system. With the advances and interest in digital libraries, there is also an increasing need of an institutional digital library facility on the intranet to encompass at one place in-house teaching and research information which can include lecture notes, question papers, preprints of publications, theses and dissertations, conference proceedings, reports of R & D and consultancy projects undertaken, etc. Once this is operational, patron interests can be well connected and linked to generate personalized alerting and selective retrieval.

The personalized system we are developing, the Selective Access Information System (SAI), for personalizing electronic resources for the faculty and researchers of the Institute has the following basic components:

- web server to host service using web tools,
- database engine MySQL to collect and manipulate user interests
- php scripts to communicate between the database engine and the web server (client request)
- client access needs only a web browser

The person has to login to the system (encrypted and stored in the database for online validation) to access his/her personalized page. The system will have provisions to gather the subject interests of researchers from their websites and the ongoing interactions they had with the library and the SAI. There is provision for the person to add/remove subject terms and various virtual links to the different services offered. These terms were used to automatically scan the different online services and e-journals available to flash a list of information resources matching the search term. There is provision for the person to select a term for a particular service such as IEL online or ScienceDirect or use more than one term on

one service or a group of services. Apart from interfacing the locally available (full text and bibliographic databases, e-books) and network accessible electronic resources (e-journals and bibliographic/full text databases), the system also provides virtual links to the following:

- Free electronic journals available on the Internet
- Free e-books on web
- News sites covering S&T as well as general
- Academic information on Internet such as websites of research and educational institutions

The user has to simply configure the choices on the personalization page to achieve suitable results. Apart from mere add/remove the user can create fresh links and sub links to reflect the different requirements. The system also provides options for searching the library OPAC, to know the borrowing details, and web form to query the librarian. A typical page for a researcher is shown in figure 1.
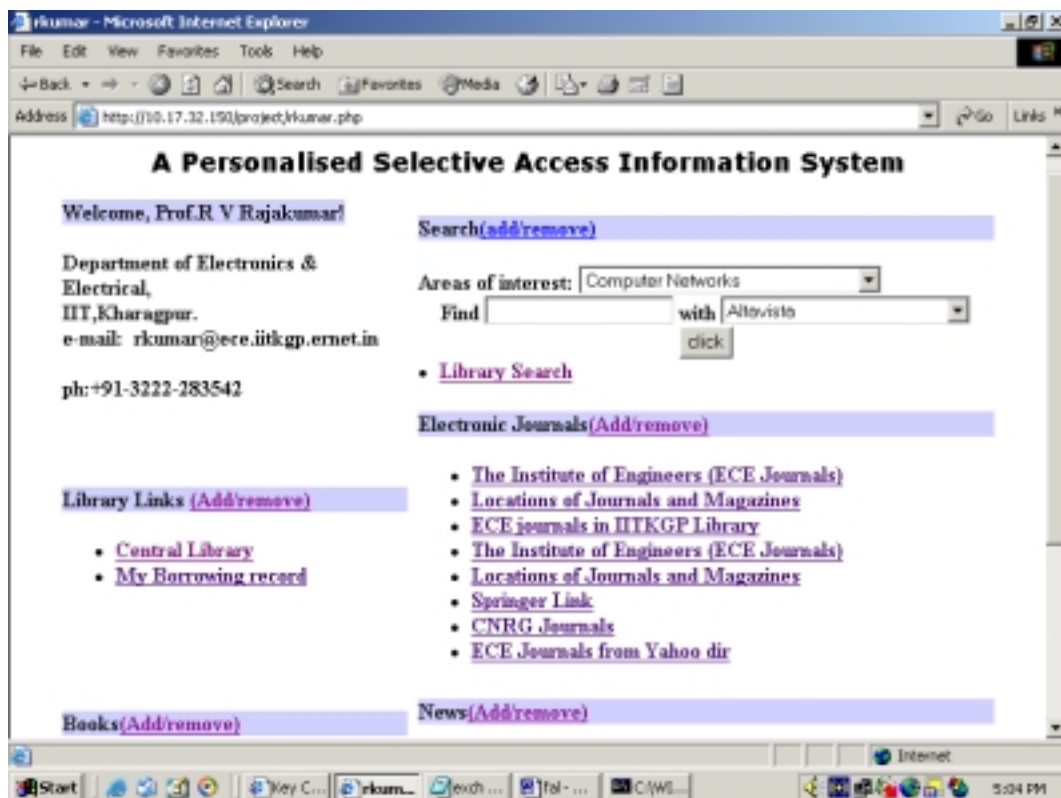

Figure 1: Screenshot of a Researcher's Personalized Page

3.   Standardizing Subject Terms

Due the lack of sudden availability of thesauri such as INSPEC, and the need for not very exhaustive subject terms to represent subject specializations of researchers in a standard manner, we looked for alternate but credible sources such as the special chapters or special interest groups of scientific and engineering societies, freely available classification schemes such as Physics and Astronomy

Classification Scheme (PACS), course headings followed by renowned universities, etc. To have a feel of the standardization provided, we discuss briefly the standardization achieved for Computer Science, especially for faculty members and researchers in the Departments of Computer Science & Engineering, School of Information Technology, and Computer & Informatics Centre. With the aid of subject terms gathered from the specializations of the 34 Special Interest Groups (SIGs) of ACM [1] and the research areas [4], courses [12] and projects [13] of faculty members in these departments, we tried to arrive at standardized representation of subject terms for researchers in computer and allied disciplines and those having computer related specializations in other branches, as shown in table 1. In the background of subject areas listed [14] in the Dept of Physics and Meteorology of the Institute and by using the Physics and Astronomy Classification Scheme [15], some categorization is attempted for few specializations in that department as listed in Table 2.

Table 1: Standard subject terms for Computer Science

Broad Term	Narrow Terms
Computation Theory/Theoretical Computer Science	complexity theory, parallel computation, computational geometry, number theory, graph theory, distributed computation, distributed processing, parallel processing, mathematical modeling, computational problem solving, symbolic manipulation, algebraic manipulation, mathematical computation, symbolic computation, data structures, simulation, modeling, cybernetics, genetic optimization, geometric modeling, high level synthesis, real time systems, number theoretic computation, algebraic computation, computational complexity, switching theory, logic design, discrete structures, automata theory, optimization techniques, operation research, computational structure, numerical analysis, chaos, fractals, dynamical systems, formal systems, fuzzy sets, queuing theory, state event systems
VLSI design	VLSI CAD, VLSI circuits and systems, VLSI design, VLSI testing
Algorithms	algorithm design, algorithm analysis, parallel algorithms
Bioinformatics	computational biology
Programming Language	formal languages, concepts, tools, design, implementation, use, specific programming languages-Ada, APL, C, OOP, visual programming
Software Engineering	process improvement, software development, CASE, object-oriented methods, software maintenance, software testing, fault tolerance, formal verification, system software, application software, software reliability
Applied Computing	experimental computing, strategic research, computing management, technology transfer
Computer Architecture	computer organization, computer design, hardware, hardware circuits, microprocessor, electrical circuits, linear circuits, pulse circuits, digital circuits, systems programming, instrumentation, modern microprocessors, optical electronics, hardware reliability, groupware
Compiler Design	compiler design, memory compiler
Artificial Intelligence	expert systems, decision support systems, machine learning, knowledge based systems, intelligent information systems, logic, logic programming, pattern recognition, machine intelligence,

	intelligent systems, knowledge acquisition, classifier learning, intelligent agents, digital circuit reasoning, speech synthesis, combinatorial optimization, knowledge discovery
Embedded Systems	design technology, hardware frameworks, software solutions
Computers-Physically Handicapped	handicapped, physical disabilities, disability solutions
Computers - Society	ethics, societal concerns, social impact, internet privacy
Computer-Human Interaction	interaction process, user interface
Data Communication/ Computer Networks	network architecture, network protocols, distributed systems, interconnection networks, LANs , telematics, telemedicine
Computer Science/IT Education	curricula, courses, syllabi, laboratories, teaching, pedagogy, software education
Design Automation	electronic design automation, computer-aided design, computer-aided testing, industrial electronics, manufacturing systems, optical electronics, robotics
Documentation	online help systems, user manual, user support
Electronic Commerce	Payment systems
Computer Graphics	image processing, robust geometric computing , exact geometric computing, digital image processing, digital signal processing, computer vision, medical imaging
Information Retrieval	information storage, retrieval, dissemination, search engines, text analysis, information systems design, information processing, information theory, coding techniques, text processing
Measurement and Evaluation	system performance, evaluation tools
Microarchitecture	instruction-level parallelism, microarchitects, microprogrammers, superscalar, computer microarchitecture, pipelined, MIMD computer architecture
Management Information Systems	business information technology
Multimedia	multimedia computing, multimedia communication, multimedia systems, multimedia storage, multimedia applications
Mobile computing	portable computers, wireless networks, nomadic computing, mobile user services, mobile networking, mobile agents, mobile communication, mobile ad hoc networks
Data management	database systems, database technology, database management systems, database engineering
Operating Systems	OS design
Security	computer security, communications security, protection, firewall, encryption, decryption, cryptographic protocols, cryptography
Internet	Web, Hypertext, Hypermedia, Internet protocols, TCP/IP

The database table for the subject term assignment contains a subject code, broad term and narrower terms and the user can enable or disable narrower terms in a broad term and when a user opts only for a narrower term, s/he may be suggested about other related terms for inclusion. Whenever a new term not in the database is suggested, it will be checked for synonyms and if not at all present, it will be added appropriately under a relevant broad term or if no suitable entries are available, it will be added as a new broad term. Presently this is done manually as it involves decision-making and requires consultation

with subject experts. However, once the database is populated with terms beyond a certain threshold, we may build up term relations into the system so that the system will be able to make suitable choices about arrangement of terms, at least up to some extent. We plan to adopt a similar strategy for other departments also and the subject terms arrived at for different departments will be merged into one database table to further the interdisciplinary and cross specialization aspects so common in technology research.

Table 2: Standard subject terms for Physics (Partial List)

Broad Term	Narrow Terms
Astrophysics	Nuclear astrophysics, Elementary particle processes, Atomic processes, atomic interactions, Molecular processes, molecular interactions, chemical processes, chemical interactions, polarization, scattering, atomic spectra, molecular spectra, hydrodynamics, magnetohydrodynamics, plasmas, relativity, gravitation, thermodynamic processes, dust processes, dark matter, cosmic rays, interplanetary space, solar physics, stellar characteristics, stellar properties, interstellar, galaxies, extragalactic, cosmology
Atmospheric Physics	Atmospheric turbulence, atmospheric gases, spectral absorption, crystal phenomena, aerosols, air pollution, scattering, polarization, image transmission, image formation, remote sensing, LIDAR, atmospheric sound, atmospheric electricity, atmospheric effects, meteorology, gravity waves, boundary layer, ionic interactions, cloud physics, climatology, weather, climate dynamics
Atomic Physics	Electronic structure, electronic transitions, electronic binding, mathematical calculations, electron correlation, electronic structure, atomic properties, atomic masses, mass spectra, abundances, isotopes, electric moments, magnetic moments, atomic spectra, fluorescence, phosphorescence, Zeeman effect, Stark effect, photon interactions
Condensed Matter Physics	x-ray diffraction, x-ray scattering, neutron diffraction, neutron scattering, electron diffraction, electron scattering, liquid crystals, crystallography, crystal defects, dislocations, radiation damage, mechanical properties, acoustical properties, lattice dynamics, phonons, phase equilibria, phase transitions, thermal properties, transport properties, quantum tunneling, quantum fluids, quantum solids, thin films, electronic structure, disordered solids, liquid metals, semiconductors, electronic transitions, electronic transport, insulators, conductivity, photons, photovoltaic effects, photoconduction, acoustoelectric, excitons, magnetoacoustic, electron states, surface conductivity, quantum Hall effect, superconductivity, magnetic properties, diamagnetism, paramagnetism, superparamagnetism, magnetic resonance, Mossbauer effect, electron paramagnetic resonance, nuclear magnetic resonance, dielectrics, piezoelectrics, ferroelectrics, optical properties, photoluminescence, electron emission, ion emission

4.   Challenges and Benefits

The prime challenge in the development of the system is to attract and sustain the user interest as researchers are too engrossed in their teaching and research activities that they hardly get time to interact with library personnel to understand and represent their specialization profiles to the desired extent. The information infrastructure in IITs are definitely of a high standard and with the spread and acceptance of electronic resources, users are virtually bypassing the libraries in many of their information requirements and they always bank upon the expertise of their friends and peers rather than approaching the library staff for assistance. We need to further improve the friendly environment so that any hesitation on the part of users in approaching the library staff will be eliminated to the maximum to share their information problems not only to improve the services offered but also to plan new services. Users must realize that the one time investment they pay in creating a better and lively representation of their subject interests will benefit them in the times to come as a result of frequent outputs from the system. In the context of a higher academic center like IIT, the subject interests of researchers keep on changing and the system has to be alert enough to gather and map the constantly changing current interests of researchers to make the system more attractive and current to the users.

Standardized subject term assignment for making the personalization system more precise and relevant in retrieval requires a special keyword assignment system which requires considerable effort in compiling the subject terms in the first instance and then grouping them as broader and narrower and synonyms require the support of a subject expert. However, the experiences of doing the task for few departments provided us with a fresh insight to undertake this task for other departments too. Since the system developed is an exhaustive list of subject interests of the important users in the Institute, this system could be used for deciding about other services such as addition of new books, subscribing new journals, etc., and also give vital inputs about making future plans and services. The system would also help to identify experts for impending research and consultancy projects.

Subject heading lists and thesaurus are not available for some disciplines whereas there are multiple sources which resort to different classification are available for some fields. Thus identification and selection of subject terms often become complex and have to be very dynamical to reflect the present research activity in the discipline and the Institute to aid and assist suitable retrievals in current areas.

As intelligence and self-learning are a part of the personalization system, we want to develop a subject term classification and retrieval system which would learn and represent new terms based on a well described heuristics and knowledge representation system. Many attributes about the users and use are residing in the client machines and the success of personalization tasks depends to a large extent on client side personalization. These are very essential to learn and represent about the user's interests, what are the contents actually used, etc., which involve new development challenges.

5.   Conclusion

Personalized services are very appealing to users as they offer seamless access to a wide set of resources available in/accessible through the library. Creating personalized customer services are projected as essential for survival in the case of service industries and personalization is projected as the next biggest challenge after digital libraries in the context of libraries. Libraries and library professionals accord utmost preference to selective search and retrieval of information resources as the library organization merges the individual requests gathered for collection development or access management and again at the time of service, these diverse interests are to be effectively identified. In the case of higher education and research, especially in the disciplines of science and technology, the efficacy of subject retrieval is time tested and thus any attempt to improve the quality of subject terms used in the personalization system enhances the utility of the system and the precision of retrievals. Libraries have to look for various sources to collect standard terms of relevance to the subject interests of users and

these are to be suitably related, carefully linked, and functionally merged into a subject term control and assignment system, though this task is challenging but crucial for the success and propagation of personalization systems.

6.   References

1. ACM Special Interest Groups Guide, http://www.acm.org/sigs/guide98.html

2. http://otn.oracle.com/tech/java/oc4j/doc_library/902/tagut/personal.htm

3. http://www-4.ibm.com/software/developer/library/personalization/

4. Research Areas, Dept of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, http://www.iitkgp.ernet.in/departments/research.php?deptcode= CS

5. http://www.rediff.com/cms/registration.jsp

6. http://my.yahoo.com/?myHome

7. http://www.amazon.com

8. Morgan, Eric Lease, Making Information Easier to Find with MyLibrary, http://www.nd.edu/~ladvance/access/issues/2002/Fall/MyLibrary.htm

9. MyLibrary@NCState, http://my.lib.ncsu.edu/

10. Builder, http://www.builder.bham.ac.uk/hld/passportmenu.asp

11. http://library.open.ac.uk/aboutus/myolib/index.htm

12. Courses, Dept of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, http://www.iitkgp.ernet.in/~cse/courses.html

13. Research Projects, Dept of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, http://www.iitkgp.ernet.in/~cse/researchmain.html

14. Research Areas, Dept of Physics & Meteorology, Indian Institute of Technology, Kharagpur, http://www.iitkgp.ernet.in/departments/home.php?deptcode=MP Physics & Meteorology

15. Physics and Astronomy Classification Scheme (PACS), prepared by the American Institute of Physics (AIP) in collaboration with certain other members of the International Council on Scientific and Technical Information (ICSTI), http://www.aip.org/pacs/

About Authors

Mr. V K J Jeevan is Assistant Librarian in the Central Library of Indian Institute of Technology, Kharagpur-721302, India. He received the Young Information Scientist Award by the Society for Information Science (SIS), New Delhi in 2000, the Second Prize in the promotional programme 'Award for Contributing Article' of Raja Rammohun Roy Library Foundation, Kolkata in 2001/2, and the PV Vergese Best Paper Award from ILA in 2002/3.
E-Mail : vkjj@rediffmail.com

Dr. Pitambara Padhi is Senior Reader at DLIS, Utkal University, Bhubaneswar – 751 004, India. He holds M.A., MLIS, LLB and Ph.D. He is Director of Academic Staff College, Counselor and Programme in charge of MLIS program of IGNOU and guided 2 Ph.Ds. He is Member in ILA, IATLIS, SIS AND ALSD and organised 4 national conferences, conventions including CALIBER-98, meetings, etc, attended more than 75 conferences/conventions/workshops, etc,. published more than 45 research papers in journals/conferences, etc, two books and edited of 3 national Conference proceedings.
E-mail: drppadhi@yahoo.com