Information and Knowledge Management Using GNOWSYS.

Meena Kharatmal, Sandhya R. and Nagarjuna G.

Homi Bhabha Centre for Science Education
(Tata Institute of Fundamental Research)
V. N. Purav Marg, Near Anushakti Nagar Bus Depot,
Mankhurd, Mumbai 400 088.

(meena, spk, nagarjun)@hbcse.tifr.res.in

Abstract

GNOWSYS (Gnowledge Networking and Organizing System), a generic system for semantic computing, is a free-software developed at Homi Bhabha Centre for Science Education (Tata Institute of Fundamental Research), and is released as an official GNU project. It is specially made for publishing vocabularies, propositions, ontologies, complex systems, web services, semantic computing and such informative resources including library resources on the web. It is designed keeping in mind the recent developments in semantic computing and knowledge representation. It is a hybrid database system with distributed, hierarchical, object-oriented, relational database support. It can serve as an information grid, it can export data in regular semantic web standards such as OWL, XTM, and its native GNOWML. It can be used for creating digital encyclopedia, thesauri, dictionaries, glossaries, multi-lingual databases, for building e-learning applications, etc. A unique feature of GNOWSYS is its potential to publish knowledge at the ultimate granular level (terms and predicates of propositions). In this paper, we shall demonstrate a case study of using GNOWSYS for library management and for knowledge management.

Keywords: Knowledge management, Information management, Digital Library, Knowledge Organization, Semantic Web.

1 Introduction

One of the ways of information communication is through the World Wide Web. The success of World Wide Web produced lots of information. Though World Wide Web is well-known today, often it is not utilized at its fullest. Usually, accessing the specific information is a trenous effort on the user as there are problems to manage the information. Inorder to overcome this problem, and to harvest the World Wide Web, new technologies need to be developed and GNOWSYS is one such tool. The paper introduces GNOWSYS (Gnowledge Networking and Organizing System) to the community and demonstrate its use for managing library, information and knowledge management. We shall present a case study on using GNOWSYS for creating the web version of the library OPAC and knowledge management.

2 Semantic Web

The World Wide Web is being used today typically for seeking information, searching for and getting in touch with other people, establishing peer groups, reviewing of research, expanding the business, reaching out globally, working on-line etc. In the Web technology, the main valuable and indispensable tools are search engines without which the Web would not have been a huge success. However, there exists serious problems associated with their use. The problems being high recall, low precision; sensitivity of results to vocabulary which occurs due to usage of different terminology, retrieval in the form of single Web page and to extract the information and collate it together calls for manual work[9].

The obstacles can be overcome by making the Web content into a machine-accessible information. The solution seems to lie on developing increasingly sophisticated techniques based on KR and computational linguistics, in order to represent Web content in a form that is easily machine-processable and to make use of intelligent techniques. This plan of revolutionizing the Web is referred to as the Semantic Web, evolving gradually out of the existing Web.

In order to create Semantic Web advanced knowledge management systems are required. The most important aim of the Semantic Web is to orga-

nize knowledge in conceptual spaces according to its meaning. The inconsistencies will be checked and new knowledge will be extracted which can be managed by the automated tools. A keyword-based search will be replaced by query answering which can be handled over several documents i.e. a queried or a requested knowledge will be retrieved, extracted and presented in a human-friendly way[9]. The vision of the Semantic Web was articulated by Tim Berners-Lee in the following way:

> The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation. — Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, Scientific American, May 2001

The Semantic Web provides a common framework that allows data to be shared and reused across applications. It is an approach to integrate data across the Web and intends to create a universal medium for information exchange by using the semantics (meaning) in a machine-understandable language to the Web contents. GNOWSYS is one such tool developed applying the semantic principles.

3 GNOWSYS: Gnowledge Networking and Organizing SYStem

GNOWSYS is an object oriented semantic engine developed using ZOPE and Python. In GNOWSYS each object is provided by an unique URL. Since it is built using a web-application server, the data can remain anywhere on the Internet, keeping only the metadata in the systems. It is possible to manage, publish and distribute knowledge by applying the semantic web principles. GNOWSYS can be used for developing ontologies, databases, electronic encyclopedias, portal sites, concept bases (knowledge bases), expert systems etc, and in general can be used for any knowledge and information management and knowledge representation. It is envisaged that GNOWSYS will help to build a semantic grid for making multi-lingual distributed concept base encompassing all knowledge of human beings, including that of science. A portal, www.gnowledge.org[1], is being developed demonstrating the said potentials of GNOWSYS.

3.1 A Hybrid Knowledge Base

GNOWSYS, is a hybrid database system with hierarchical, relational, distributed, object-oriented, database support. In a hierarchical database, records are grouped in a logical hierarchy, connected in a branching structure. A relational database is a database constituting of set of relations in the form of tables in which the records are in the form of attribute values. In the object-oriented database, the contents are stored in the form of objects, with features of inheritance (the values which are assigned for parent objects are inherited to the child objects assigned), polymorphism (it is the ability of objects belonging to different types to respond to methods of the same name). A distributed database is a database that is under the control of a central database management system in which storage devices are not attached to a common CPU. It may be stored in multiple computers located in the same physical location, or may be dispersed over a network of interconnected computers[6].

3.2 Architecture

The architecture of GNOWSYS, is designed keeping in mind the recent techniques involved in knowledge representation (KR) area. KR in artificial intelligence (AI) is used to study formalisms to model human knowledge and for problem solving. GNOWSYS can be applied for drawing concept graphs, semantic nets, concept maps, SemNets which are currently used by several researchers and educationists[20, 19, 11] to enhance conceptual learning in the context of science education. The architecture of GNOWSYS is structured to accommodate different dimensions of KR such as—*generality, semantics, complexity, inference*[12] as shown in Figure 1.

We come across a wide variety of concepts in the form of particulars or generals or even abstract. Inorder to understand the wide variety of concepts in our discourse it is essential to organize the concepts based on their order of generality. Along the *generality dimension* of GNOWSYS, the three different levels of generality such as tokens (particulars), types (generals), and metatypes (types of types), can be organized. For example, it is possible to organize knowledge of the taxonomical classification system along the generality dimension. In the taxonomical knowledge base, there exists concepts such as—mammals, dogs, cats, Fido, Tom, humans, James Watson, etc. The generality dimension is applied when organizing the above

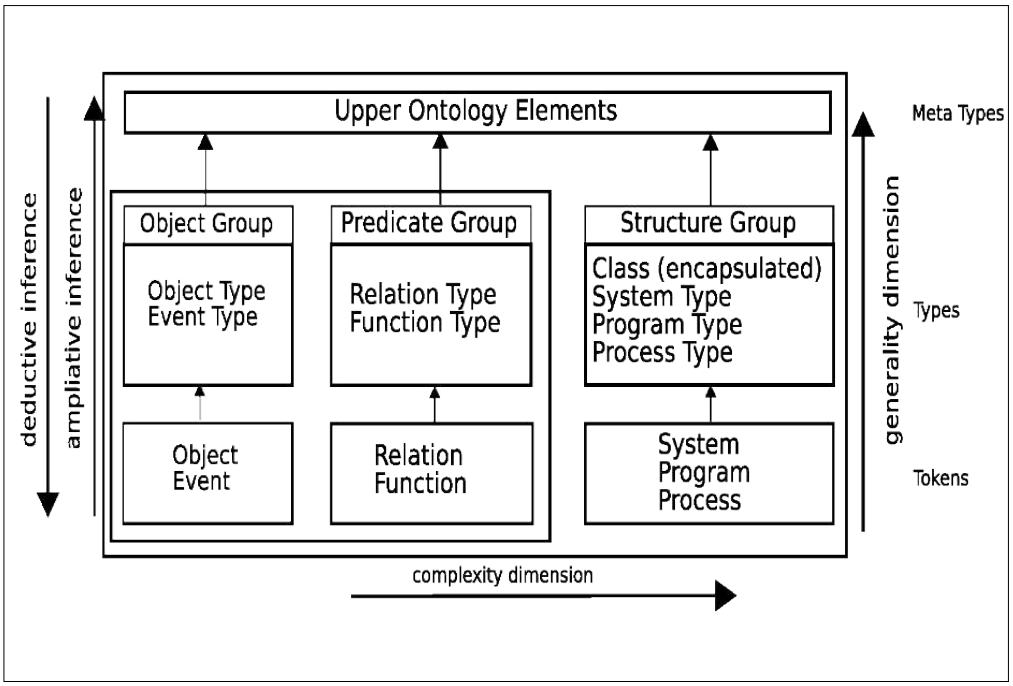


Figure 1: A diagram representing the architecture of GNOWSYS.

knowledge base—Fido, Tom, James Watson are the particulars and hence organized in the token layer; dogs, cats, humans are generals and hence are organized in the type layer; and taxonomical concept, person concept are organized in the metatype layer.

The semantics and consistency check is carried out along the *semantic dimension* of GNOWSYS. To start with, in the first layer all kinds of propositions are allowed to store without any semantic constraints. This is observed in the case of novices when they start with representing their knowledge in idiosyncratic way without following any constraints or principles. The propositions are stored in the form of *well formed formulae* (WFF). In the second layer, the WFFs can be combined with the semantic constraints, logical connectives, modalities, propositional attitudes, quanti-

fiers etc. The consistency is implicit at this layer and hence it is referred as *implicit structured system* (ISS). In the third layer, the validity constraints can be imposed explicitly and therefore gives rise to *explicit consistent system* (ECS) which is quite similar to the experts' knowledge system. The semantic dimension allows to represent novice's and expert's knowledge (alongwith the transformation), with the semantic constraints and validity. The semantic dimension is implicit throughout the entire knowledge base built using GNOWSYS, and hence it is not represented in the figure.

The basic components of a knowledge base are ObjectType (OT), Object (O), RelationType (RT), Relation (R), MetaType (MT), EventType (ET), Event (E), FlowType (FT), Flow (F). These components help to store the terms, propositions and procedures. The *complexity dimension* helps to generate complex compositions using the structure groups consisting of ProcessType (PT), Process (P), StructureType (ST), Structure (S), Encapsulated Class, Programs and ProgramType.

Using the *inference dimension* of GNOWSYS, the epistemic values such as validity and truth can be checked. At present, there does not exist any in-built module supporting for inferences but with developing an interface between the GNOWSYS and any existing inference engines, it is possible to deduce consequences using deductive inference.

3.3 Underlying Technology

GNOWSYS is developed as a product of ZOPE, by extending ZOPE classes and using Python programming language. ZOPE (Zee Object Publishing Environment)[7] is a free (as in freedom) web application server for building content management systems, intranets, portals, and custom applications. ZOPE allows to store content and custom data, dynamic HTML templates, scripts, catalogue, and connections to relational databases. It features a strong through-the-web development model, allowing to update the website from anywhere in the world alongwith powerful integrated security model. It is possible for connecting not only to ZOPE's object database, but also with relational database allowing for strong data integrity. It runs on the GNU/LINUX, UNIX-based platforms as well as Windows NT. It can be used with most popular web servers or its own built in web server.

ZOPE is designed for Web object as well as Web development model. It is written using Python, an interpreted, interactive, object-oriented programming language with clear syntax and dynamic semantics. Python[4] emphasizes readability due to its clear syntax and therefore is easy to maintain. GNOWSYS is interoperable, i.e. it works on all the known operating systems.

3.4 Semantic Computing

As presented above, GNOWSYS has procedural objects, alongwith declarative objects, such as function, program, and class. GNOWSYS is unique in this feature that it can store and activate (executable) objects. Using this feature, it is possible to design applications without writing programs in any programming language: i.e., specifying the semantics of a program and mapping the elements of the program to the surrogates of procedures is sufficient for GNOWSYS to test the application design. This forms the foundation for semantic computing using GNOWSYS. This feature enables anyone, even those who do not have computer science or programming background, to develop a knowledge base provided they can grasp the logic of the domain.

4 Dynaminc Information Management Using GNOWSYS

In the present scenario, with the help of available library software packages, it is possible to organize the library resources. But this is not enough, because the resources are not always books and journals, there is also more information available on the internet. If we want to fully exploit or utilize the information available globally, we need to go beyond just organizing the library collection to information and knowledge management. The existing softwares do not meet the above requirements, and hence GNOWSYS plays a major role in managing all kinds of digital resources alongwith traditional resources. Using GNOWSYS it is possible to integrate the functionalities of information management such as managing the documents and also for planning and coordination.

Information management involves anything from library management, to office management, or any types of record management. Due to bud-

get constraints, most organizations find difficult to implement IT for their services including libraries. It is important for managing databases, office resources, providing online access over the internet. To cater to all these requirements, usually there are special applications for each purpose. A generic information management system like GNOWSYS can serve all the requirements in the office such as accounting, creating inventories, managing leave records, etc., and also for replacing the specialized software.

The reader would think that all the above requirements can be achieved using any database, and would also like to know what is so special about GNOWSYS. We would like to emphasize that GNOWSYS is especially developed for managing dynamic knowledge base systems. It is made to model the cognitive development (or the development of human knowledge). For example, in a human's ontogeny (lifetime), knowledge is acquired by changes in conceptual schemes, and it gets developed through various processes. It is very essential to capture these changing conceptual schemes as they get developed. KR is one such alternative for modeling human knowledge, and GNOWSYS which is based on KR principles, allows for representing various conceptual schemes in one's ontogeny or learning course. GNOWSYS allows one to begin representing the knowledge starting from loosely formed structures which can at the same time have contradictory beliefs, semantic inconsistencies in an implicit form and over a period of time transforming the knowledge into an explicit and consistent form. This is what we assume in the context of learning. In this way GNOWSYS can capture the development of human knowledge because it always preserves the earlier knowledge structure, a feature which is not found in the conventional database management system (DBMS) due to their rigid structures or schemes.

In the conventional DBMS, in order to create database the scheme should be frozen before adding the data. However this does not allow to incorporate changes in the existing schemes, and one ends up in creating different schemes as and when required. In this way the schemes in DBMS are rigid. On the contrary, GNOWSYS is flexible because its schemes can change according to the input and the requirements. The schemes in GNOWSYS are generated as they get developed while we input the data. GNOWSYS helps to represent developing systems or changes.

5 Library Management Using GNOWSYS

The reader may have realized that GNOWSYS can be used to create wide variety of applications related to information management. In this section our focus is to present a case study on using GNOWSYS for efficient library management system. Based on the three layer architecture of the GNOWSYS, we organized the metadata of the library. Metadata is data that describes data or content objects. The metadata was categorized in the form of ObjectType, Object, AttributeType, Attribute, RelationType, Relation. Objects are classified according to subclass/superclass relation, and also instantiation relation (class-instance). Relations are assigned between objects and their classes (called ObjectTypes or Metatypes). The library's metadata includes ISBN, author, editor, publisher, author mark, pages, subject, class number, accession number, etc. The ObjectType is used for classification. The subject is classified as ObjectType since it is a class. The unique feature of a book or a journal for example, accession number or periodical number are assigned as ID, and those such as ISBN, ISSN, are classified as AttributeType which can be assigned values. Every physically existing material i.e. each book or each journal is classified as an Object. The organization of the library metadata is shown in figure 2.

When creating entries for each book or each journal, one has to select its parent ObjectType which is nothing but the subject to which it belongs. Once the object is created, then values are to be assigned for each attribute of the book.

5.1 Library on the World Wide Web

The Centre's library database is currently managing the database of around seventeen thousand resources on the existing proprietary software. Using GNOWSYS, this library database was launched on the World Wide Web. Since GNOWSYS is a web-application server, incidentally also developed in-house, the library database can be uploaded on the internet. Inorder to create a web version of the library collection, we worked on the data exchange model. First the metadata of library information science (LIS) was translated into GNOWSYS model. The database (.dbf) file of the existing software was first opened in a spreadsheet and saved as tab delimited text file. Using Python, a script was written to read each and every field of the

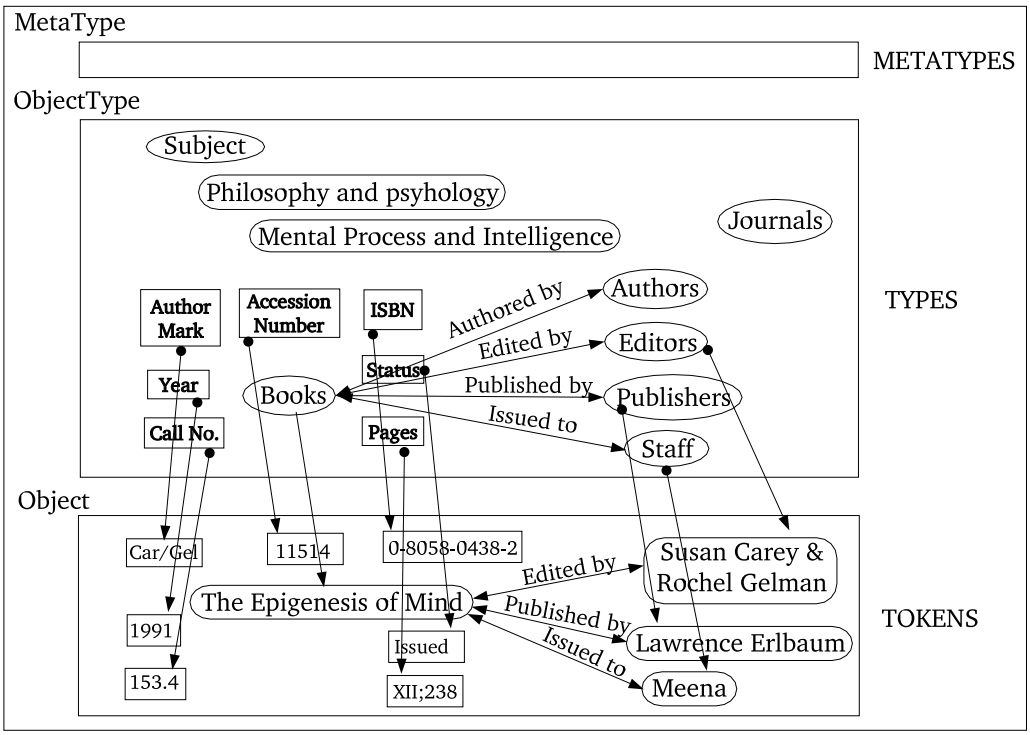


Figure 2: The organization of the library metadata.

database tables which was created using previous database management system. After the successful generation of the script, each and every field of the database tables was fully converted into objects in GNOWSYS. Once the database was successfully imported the display on GNOWSYS was designed in the user-friendly way alongwith the two standard formats viz., MARC and CARD. One of the most widely used feature in the library is to query for the collection. The search script enables one to generate results of query based on author, editor, title, publisher, ISBN and keywords. Since both data and metadata of objects can be stored, indexed in a catalogue it enables for a faster query.

5.2 Data Exchange Using GNOWSYS

GNOWSYS supports data exchange in several forms. It can accept any import filter as long as the database exports into text or XML (eXtensible Mark-up Language). XML, a sub-set of earlier mark-up language called the Standard Generalized Markup Language (SGML), is an open standard used for defining, validating as storing structured data objects by expressing them as tagged text in Unicode. The data-exchange modules of GNOWSYS will support various standard knowledge representation schemes such as CL (common logic), PetriNets, KIF (knowledge interchange format), CG (concept graphs), OWL (web ontology language), XTM (XML Topic Maps), etc.

5.3 Managing User Profile

A staff user profile is created by importing the user's database. This facilitates the circulation as well providing selective dissemination of information (SDI), current awareness services (CAS). This allows each user (with secured login ID and password) to recommend or reserve the books online, or request for any other information. It is also possible to view the table of contents, generate bibliographical entries for citations.

5.4 Selective Dissemination of Information

Since an user profile already exists, it is much easy to create relations between the areas of interest of the users with that of the collection (i.e. keywords). Using GNOWSYS, it is possible to establish linkages or relations between the users and new arrivals in the library. Whenever there is a new arrival in the library, users are alerted. It can generate online request forms regarding abstract, review, etc. Current awareness in the form of new arrivals, contents of current journals, clippings of newspaper articles can be disemminated to the users using the user profile.

5.5 Digital Libraries

In the IT age, there is a growing demand for the latest information. The digitization of the library involves "series of activities that brings together collections, services, and people in support of the full life cycle of creation, dis-

semination, use, and preservation of data, information, and knowledge"[18]. The main advantages of a digital library are that the digitized information can be accessed from anywhere in the world, it has fast access, the storage, management and retrieval of information is precise and easy, and the digitized information can be easily converted into a hard copy. The data of the digitized information is transmitted in its exact form (images, video, audio, text, etc.). GNOWSYS can be used for digital document management system (DMS), for any large organization or for personal use. It allows for creation, storage, retrieval and dissemination of documents in electronic format.

World Wide Web made possible several digital libraries around the world for example, Wikipedia, Wordnet, OpenCyc, Netscape's Open Directory which has a storage of digital data. Using the metadata of GNOWSYS, it is possible to create relations with these digital libraries and produce a structured and organized knowledge base.

6 Content Management Using GNOWSYS

Content Management (CM) is a set of processes and technologies which can create, update, publish, translate, archive, and retire. To start with, an instance of digital content is created by authors, and later the content is edited (by editors), and on approval of the content it is ready for publication. The next generation of content management is best referred to as semantic content management, where the content carries meaning that is expressed by metadata according to a semantic structure based on open standards[22]. The semantic content management is about managing content objects based on their properties.

7 Knowledge Management Using GNOWSYS

GNOWSYS is being used as a tool for knowledge management in science education. At present, there exists wide variety of tools for knowledge representation (KR), of which the most widely used are Concept Maps[20, 19], SemNet[11], semantic network, concept circle diagrams[19], concept graphs[23]. The graphs generated from these KR tools are usually stored

separately thus making the knowledge base inaccessible for the users across the network. However, in the design of GNOWSYS, the graphs generated by various applications can be shared and published by the system through XML based representations schemes.

GNOWSYS is being used in research on "Knowledge Organization in Biology Education". A knowledge base of concepts and semantic relations is being developed using GNOWSYS[15]. Our research study is to characterize and organize knowledge based on KR using the grammar of scientific knowledge. The research methodology of organizing knowledge is based on KR approach which involves to apply logical principles. Using GNOWSYS, we have created a knowledge base of some biological terms wherein the knowledge is classified and organized into MetaType, ObjectType, RelationType, Relation, Objects, Attributes. Some of the objectives of our research are—to organize concepts based on their cognitive function (role); assign valid and authentic semantic relations (knowledge organizers) to the concepts; to compare and restructure (reorganize) the novice's knowledge structure with that of an experts knowledge structure. Using GNOWSYS, graphical representations such as concept maps, concept graphs can be generated based on the knowledge base. The knowledge organizers help eliminate ambiguity, maintain parsimony and apply precision to the scientific body of knowledge[16].

The figure 3 shows a principled concept map on organic molecules generated from our knowledge base. It depicts the knowledge organizers i.e. relation types which help to represent some of the knowledge about organic molecules.

8 Conclusion

GNOWSYS, a web application server, is developed in-house and is released an an official GNU project. It is a free (as in freedom) software i.e. it gives freedom to run the program, study the program and adapt or improve/change the program, and also to redistribute the program to the community.

The objective of the paper is limited to introducing GNOWSYS to information scientists. We have shown that GNOWSYS can be used in the

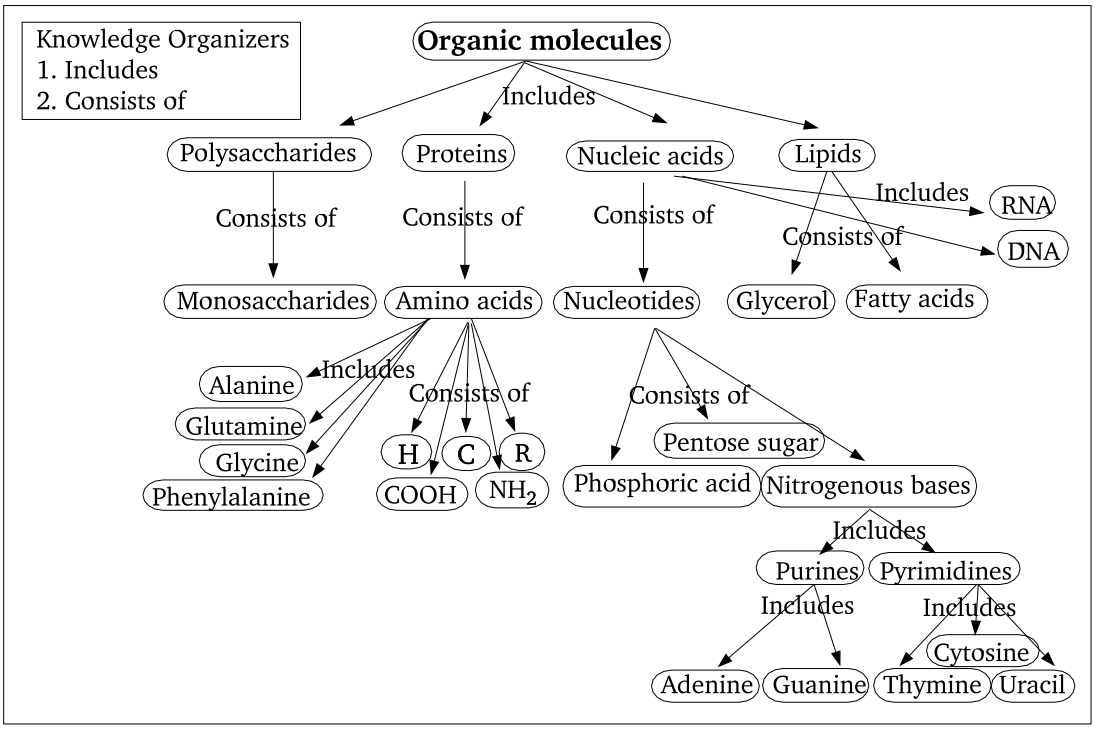


Figure 3: Knowledge organizers for understanding organic molecules.

area of information and knowledge management for content management, creating digital documents, etc.

Acknowledgements: We wish to thank Bipin Apandkar and Prasanta Baruah for providing the programming support for library management and Shrikant Gaikwad for providing the technical support.

References

[1] Gnowledge portal site. Website. http://www.gnowledge.org.

[2] Hbcse library website using gnowsys. Website. http://www.hbcse.tifr.res.in:9673/library/hbcse-library/.

[3] A multilingual thesaurus. Website. http://www.hbcse.gnowledge.org/wordnetbth/.

[4] Python. Website. http://www.python.org.

[5] Website of homi bhabha centre for science education. Website. http://www.hbcse.tifr.res.in/.

[6] Wikipedia, the free encyclopedia. Website. http://en.wikipedia.org/wiki/Database_systems.

[7] Zope (zee object publishing environment). Website. http://www.zope.org.

[8] GNOWSYS: Gnowledge networking and organizing system. Website, 2005. http://www.gnu.org/software/gnowsys/.

[9] Grigoris Antoniou and Frank van Harmelen. *A Semantic Web Primer*, chapter The Semantic Web Vision, pages 1–21. The MIT Press, Cambridge, 2004.

[10] K.S. Raghavan S.B. Ghosh N.R. Satyanarayana S.K. Savanur Muttayya Koganuramath T.J. Mathew D. Kamalavijayan, Harsha Parekh, editor. *International Conference on Information Management in a Knowledge Society*, volume I & II, Mumbai, February 2005. Indian Association of Special Libraries and Information Centres, Allied Publishers Pvt. Ltd., New Delhi.

[11] Kathleen Fisher, James Wandersee, and David Moody. *Mapping Biology Knowledge*. Kluwer Academic Publishers, The Netherlands, 2000.

[12] Nagarjuna G. GNOWSYS: A system for semantic computing. Website, 2005. http://www.gnu.org/software/gnowsys/gnowsys_files/cpv2/.

[13] Ramesh C. Gaur. *Re-engineering Library and Informational Services: Process, People & Technology*. Allied Publishers Private Limited, Mumbai, 2003.

[14] D. Kamalavijayan. *Information and Knowledge Management*. Macmillan India Ltd., Delhi, 2005.

[15] Meena Kharatmal and Nagarjuna G. Knowledge organization of biology education. Website, 2004. http://zope.hbcse.tifr.res.in:9673/Meena/Biology/.

[16] Meena Kharatmal and Nagarjuna G. Understanding science through knowledge organizers: An introduction. In *An International Conference to Review Research on Science, Technology and Mathematics Education*, Goa, India, December 2005. HBCSE, HBCSE. http://www.hbcse.tifr.res.in/episteme/themes/meena-knowledge-organizers.

[17] Doug Lenat. The cyc knowledge server. Website, 2005. http://www.cyc.com/cyc/technology/.

[18] Gail McMillan. Digital libraries support distributed education. Website, 1999. http://www.ala.org/ala/acrl/acrlevents/mcmillan99.pdf.

[19] Joel Mintzes, James Wandersee, and Joseph Novak, editors. *Teaching Science for Understanding — A Human Consctructivist View*. Academic Press, USA, 1998.

[20] Joseph Novak and D. Bob Gowin. *Learning How to Learn*. Cambridge University Press, UK, 1984.

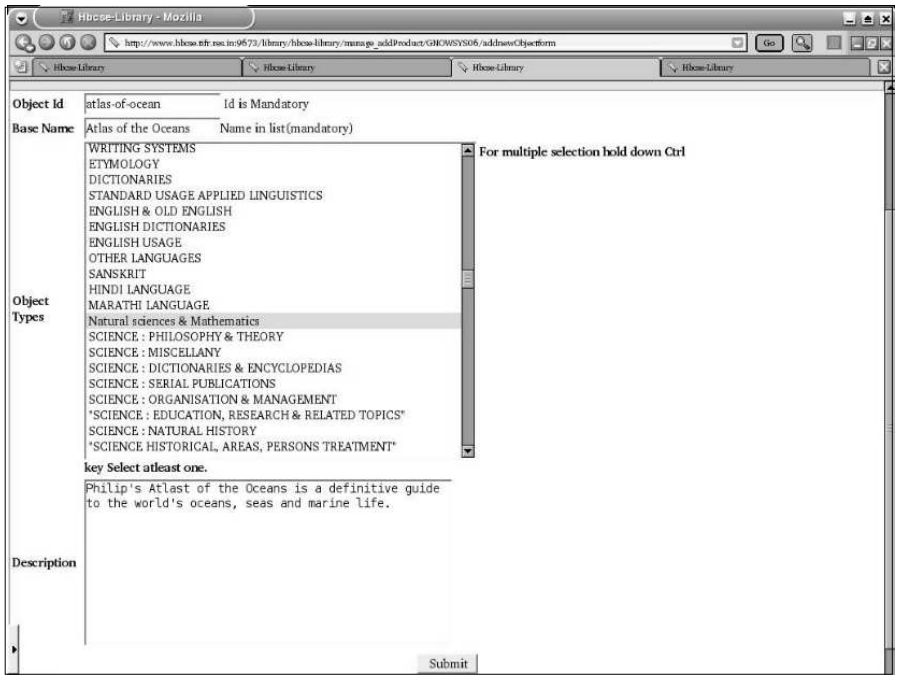
[21] National Library of Medicine. Unified medical language system (umls). Website, 2005. http://www.nlm.nih.gov/research/umls/.

[22] Profium White Paper. Technical impact of semantic on content management. Website, October 2004. http://www.profium.com/library/Profium_technical_white_paper.pdf.

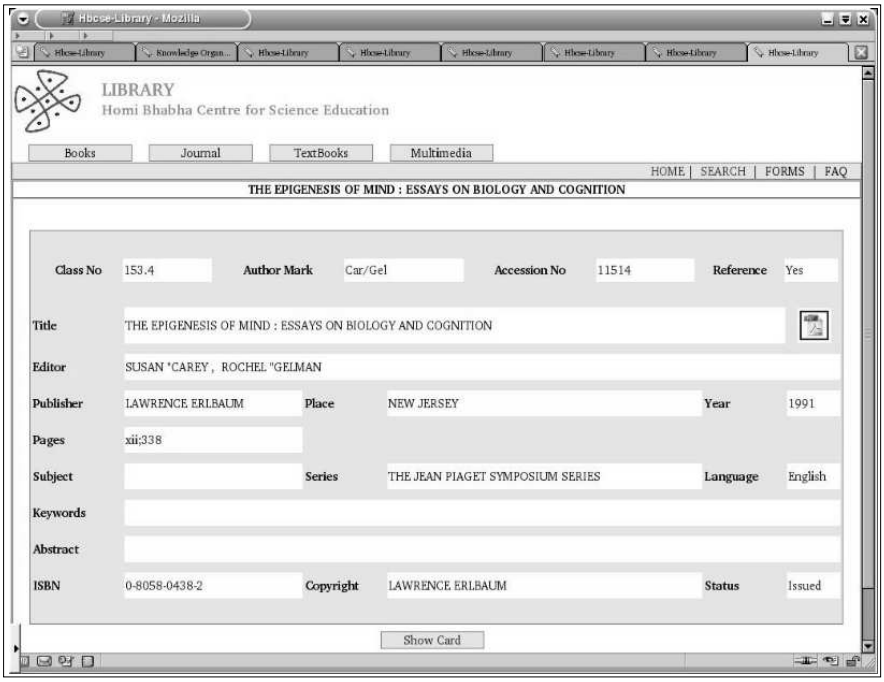
[23] John Sowa. *Knowledge Representation: Logical, Philosophical and Computational Foundations*. Brooks/Cole, USA, 2003.

[24] Morton Winston, Roger Chaffin, and Douglas Herrman. A taxonomy of part-whole relations. *Cognitive Science*, 11:417–444, 1987.

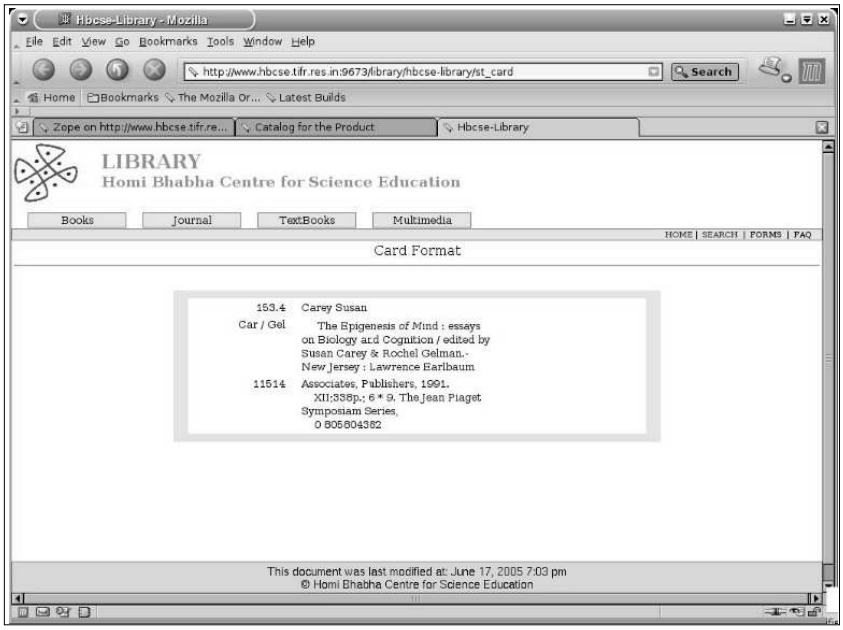
Screen Shots



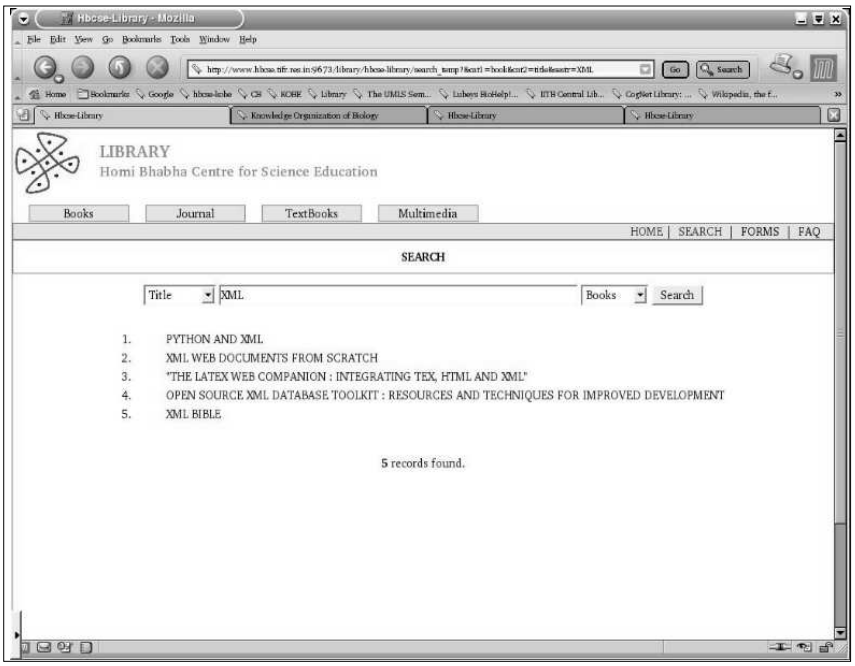
Screen shot 1: Adding an entry in library management system using GNOWSYS.



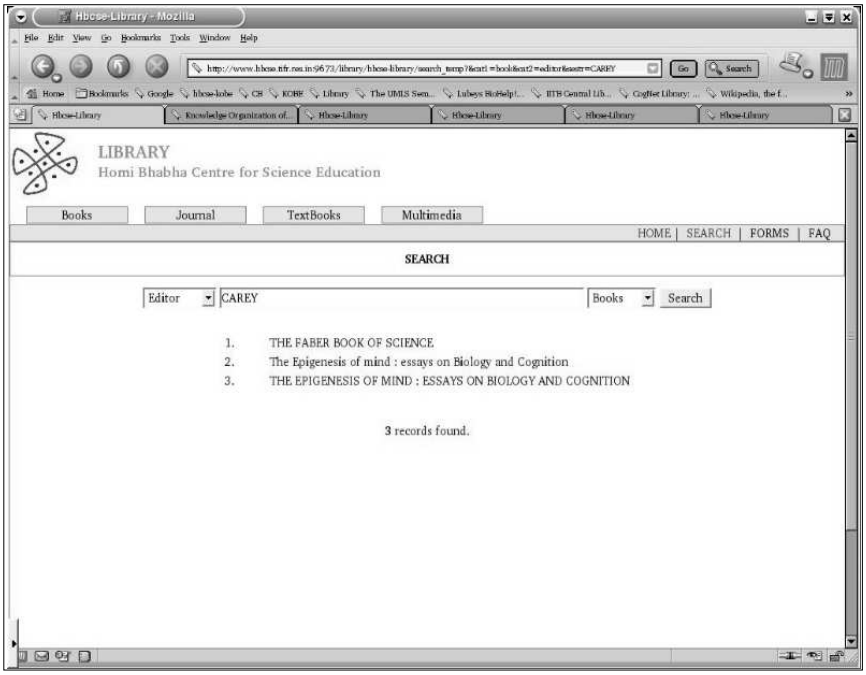
Screen Shot 2: Display of book using GNOWSYS.



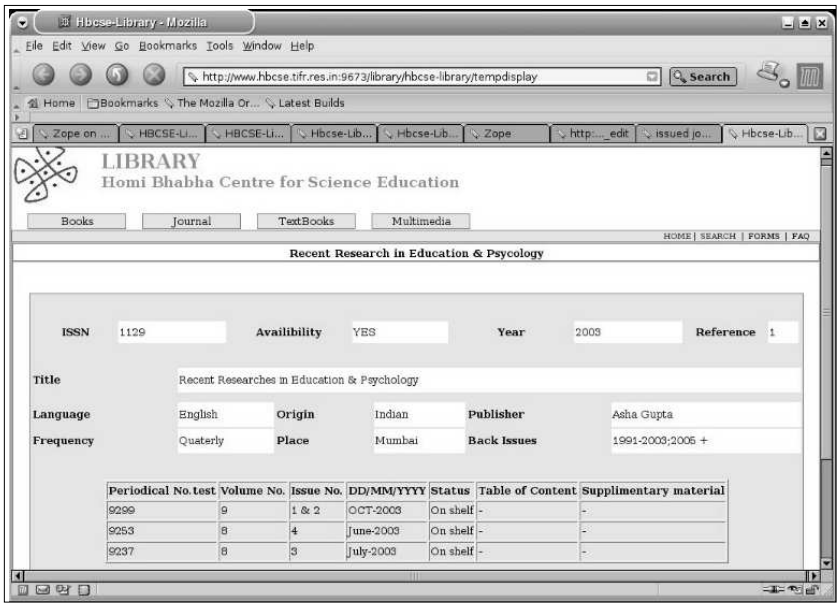
Screen Shot 3: Display of book in CARD format using GNOWSYS.



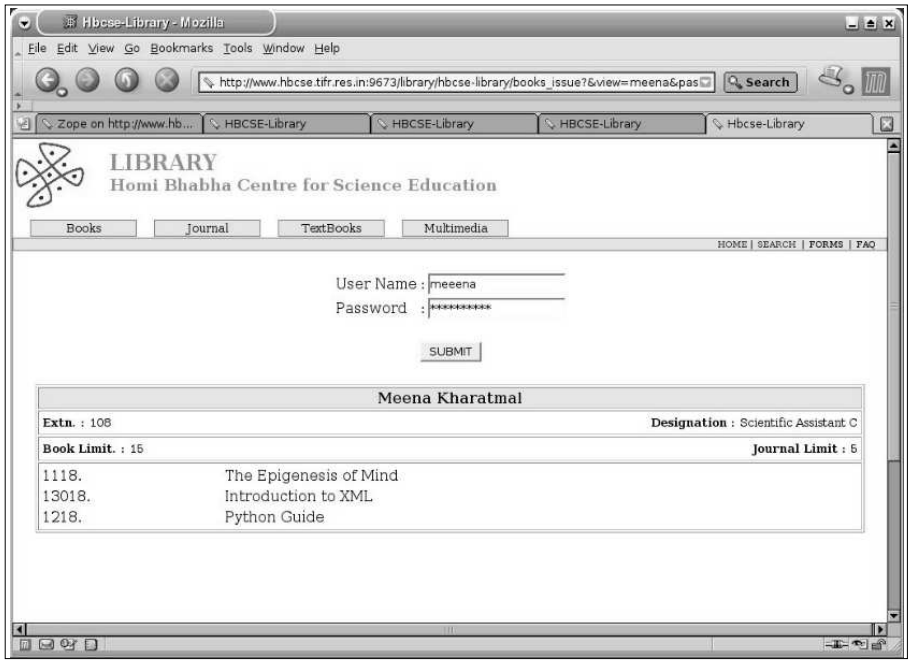
Screen Shot 4: Display of query for book based on title.



Screen Shot 5: Display of query for book based on author/editor.



Screen Shot 6: Display of the query for journal.



Screen Shot 7: Display of the user profile.

Title: Information and Knowledge Management Using GNOWSYS.

Authors: Meena Kharatmal, Sandhya R. and Nagarjuna G.

Affiliation: Homi Bhabha Centre for Science Education
(Tata Institute of Fundamental Research)

Abstract:

GNOWSYS, a generic system for semantic computing, is a free-software developed at Homi Bhabha Centre for Science Education (Tata Institute of Fundamental Research), and is released as an official GNU project. It is specially made for publishing vocabularies, propositions, ontologies, complex systems, web services, semantic computing and such informative resources including library resources on the web. It is designed keeping in mind the recent developments in semantic computing and knowledge representation. It is a hybrid database system with distributed, hierarchical, object-oriented, relational database support. It can serve as an information grid, it can export data in regular semantic web standards such as OWL, XTM, and its native GNOWML. It can be used for creating digital encyclopedia, thesauri, dictionaries, glossaries, multi-lingual databases, for building e-learning applications, etc. A unique feature of GNOWSYS is its potential to publish knowledge at the ultimate granular level (terms and predicates of propositions). In this paper, we shall demonstrate a case study of using GNOWSYS for library management and knowledge management.

Mailing Address: Homi Bhabha Centre for Science Education
V. N. Purav Marg, Near Anushakti Nagar Bus Depot,
Mankhurd, Mumbai 400088.

Telephone Number: 022-25555242 / 25554712

Fax Number: 022-25566803

Email Address: meena@hbcse.tifr.res.in