

ORGANISING DIGITAL INFORMATION USING OPEN SOURCE SOFTWARE: AN ATTEMPT FOR KNOWLEDGE MANAGEMENT IN CORPORATE SECTOR

Satish S. Munnolli
Asian Paints (I) Ltd
L.B.S. Road, Bhandup West
Mumbai 400 078, India
Email: satishmunnolli@yahoo.com

An attempt has been made to develop a database of R & D reports and trade literature using open source software – 'GreenStone Digital Library' (GSDL). The reports in digital form are generated within the organisation and the data downloaded from INTERNET are also included in the database. The paper discusses the approach to have Knowledge Management system, which organises the digital documents in a library system and its effective utilisation in Intranet environment.

INTRODUCTION

The concept of information and its format has dramatically changed over the last few years because of the revolution in Information Technology (IT). Development of new softwares, upgradation of existing softwares and emergence of new file formats are never ending phenomena. Till recent past, the role of a librarian was to collect, organise and disseminate printed documents. But, the pace at which the digital documents are now growing is of concern to library professionals as they have to organise and manage the digital information in the changing environment. Various softwares are being developed to control and manage the digital documents in networked environment.

An attempt has been made here to collect and organise the digital documents that are generated in various formats by the R&D division of Asian Paints Ltd and also the downloads from INTERNET for its effective utilisation in Intranet environment.

VALUE OF IN-HOUSE INFORMATION

Besides conventional resources such as books, journals, standards and patents, research

reports, minutes of meetings, data analysis reports and survey reports are important sources that carry a lot of vital information in any organisation for designing and developing a new product / new process or service.

Scientists working at various levels need a variety of information from outside as well as from those generated within the organisation. Apart from conventional resources available in a library, a lot of technical literature available from suppliers/vendors and other trade literature downloaded from INTERNET are referred to by the scientists.

At Asian Paints Limited, technical literature includes physical and chemical properties of raw materials, formulations, safety measures, and information related to manufacture of paint and coatings. Raw materials from hundreds of companies are available and it is definitely a huge task for the customer of raw material to choose the best amongst available ones, which has maximum properties at a minimum cost and that suits the desired end product. It is not the claim of supplier on properties of raw materials they are supplying that is considered but the sample raw materials are used and tested in various products in laboratories to know the reality on properties about its correctness. If a particular raw material passes all the tests and meet the standards (set by company, as well as national and international standards) it gets final approval for the bulk purchase for its use in designing desired product in the plants in mass scale. Since each department has its own role to play and perform to achieve the objective, every department procures the desired data in its capacity either through direct contact or through INTERNET. The entire process results

in flow of voluminous data and is scattered across the company in various departments at various levels. Series of experiments conducted in laboratories, brainstorming sessions and discussions at various levels are documented and reports are generated in the soft version before the final print is made available in hard copy. These reports are invaluable to the company for comparison of properties of similar raw materials and further research and development activities.

KNOWLEDGE MANAGEMENT

The concept 'Knowledge Management' (KM) has varied connotations for various professions. KM is defined in many ways that suits the environment of an organisation and there is no universal definition that has been accepted across the world [1]. However, in simple terms KM is 'knowing what you know and then bringing profit from it'. It is capturing the intangible knowledge of individuals and making it available in tangible or explicit format for the use of others. It facilitates in tapping the human knowledge for benefit of others in achieving organizational goals. KM differs from organisation to organisation in the way they create, capture and reuse the knowledge and few definitions are as below.

"KM is a practice of harnessing and exploiting intellectual capital to gain competitive advantage and customer commitment through efficiency, innovation and faster and more effective decision-making" [2].

"KM is the generation, representation, storage, transfer, transformation, application, embedding and protecting of organizational knowledge" [3].

"KM caters to the critical issues of organizational adaptation, survival, and competence in face of increasingly discontinuous environmental processes that seek synergistic combination of data and information processing capacity of information technologies and the creative and innovative capacity of human beings" [4].

"KM is the name of a concept in which an enterprise consciously and comprehensively

gathers, organizes, shares, and analyzes its knowledge in terms of resources, documents, and people skills" [5].

From the above definitions it can be concluded that it is a process of transforming tacit knowledge into explicit information and making it available to users for effective utilisation of resources for the benefit of organisation as a whole. KM is often facilitated and supported by IT but technology by itself is not KM. IT enables to complete the concept of KM in an organisation. It helps in creating an environment wherein the knowledge flows smoothly in an organisation.

WHY IS KM REQUIRED IN R & D?

With the pace at which information is generated, it is not possible to assess the value of every piece of information. Further, one individual's rating on importance of information differs from that of another. The KM system captures and integrates knowledge of individuals and makes it available at a common place. It connects the people involved in similar activity and encourages bonding in teams or research groups. It helps in enhancing the process of development activity and results in accelerating innovations.

A Vice President, R & D of a Fortune 500 chemical company says – "... we spent two years in developing a new technology to reduce capital expenditure by two-thirds. The key research engineer left in the middle of the project. The new engineer had to spend a whole year getting upto speed and the project took over four years to complete instead of three [6]" Such situations are a common experience in many companies.

A study by Kinematic Inc., U.S.A. [7] reveals that an average researcher spends only 11% of his time on creative activity and out of this only 2% of the time is spent for generation of original ideas (Fig 1).

Research reports are generated and are available in concerned departments only.

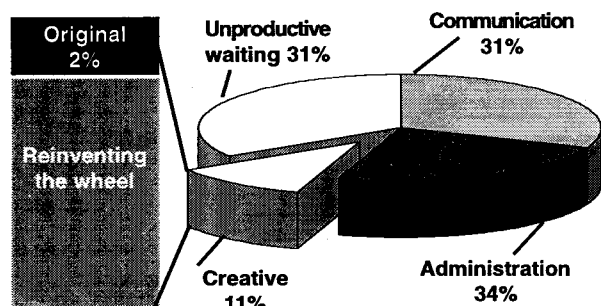


Fig. 1 — How researcher spends his time

Unless there is a request from another department, the generator of the report does not keep the report in a common place.

OBJECTIVES

Objectives of this study are:

- To create an ideal platform for sharing knowledge derived while processing and developing new products;
- To build a digital database of research reports and make them available and accessible to all scientific staff in R & D; and,
- To accelerate the innovation process in R & D by setting up a knowledge database.

NEEDS IDENTIFICATION

It is necessary to identify the needs of individuals and groups before proceeding to develop any database. Information generated in R&D organisations are important and there are many issues regarding keeping information in common place especially in the corporate sector.

Project reports / results, raw material analysis reports, literature of raw materials, raw material data sheets, product evaluation reports and reports on out door studies are identified as core data that are required by groups at various levels of research project for scientists working in the laboratories. Information such as formulation sheets are not considered in KM.

SELECTING A SUPPORTING SOFTWARE

There are several softwares available in the market for Knowledge Management. In this study GreenStone Digital Library has been selected. The reason being it is an open-source software, available from <http://greenstone.org> under the terms of the GNU General Public License. It provides a new way of organising information and publishing the same on the Internet, Intranet or on CD-ROM. GSDL is produced by the New Zealand Digital Library Project at the University of Waikato, and distributed in cooperation with UNESCO and the Human Info NGO.

GSDL runs on Windows, Unix and Mac OS X. The distribution includes ready-to-use binaries for all versions of Windows, and for Linux. It also includes complete source code for the system, which can be compiled using Microsoft C++ or gcc. GSDL works with associated software that is also freely available: the Apache Webserver and PERL. The user interface uses a Web browser: typically Netscape Navigator or Internet Explorer. GSDL is available both in local library and web library version. The major difference between the two is that the *local library* contains its own built-in webserver. The *web library* however, requires an external webserver like Apache or Microsoft IIS. In the present study, GSDL (local library) was used on Windows platform with a Personal Web Server.

DESIGN OF KM SYSTEM

After identifying the needs and selecting the suitable software, the system was designed. The system was designed to meet all user expectations at all levels and the database made accessible in an Intranet environment at all nodes. However, the controlling capabilities was made available at a single node. A typical KM process involves following four stages:

- Capture of data
- Organisation and Storage
- Distribution, or better sharing
- Application or Leverage

It is always useful if the input data is in standard format across the various departments. But, reports are not bound to be so. If one goes for maintaining uniformity in reports for the suitability of database, it beats the actual purpose of research or developmental activity. It is also equally difficult to bring or convert all the reports in a standard format because it is labourious and time-consuming process. So without giving much emphasis on format, the reports are captured for developing database because GSDL takes care of all these aspects while retrieving any document from the built database.

BUILDING DATABASES

Building a database of reports is a team effort and the process involves sharing information by individuals on a common platform. It is the most difficult stage in establishing a KM system. In the beginning lot of inhibitions were shown by contributors in sharing their reports with others.

The nature of activities and outputs, the boundaries of groups, its values, and organisation goals were briefed to participant contributor in developing a healthy community. The importance of reports generated by individuals and their role for the development of organisation was convinced. Voluntary contribution, downplaying the role of individual powers establishes growth of knowledge sharing amongst groups and it creates new relationships in practice.

In every department, instructions were given to create a separate folder in the departmental computer and to copy all the reports in that folder that are of interest to other groups and are sharable. Simultaneously in the library computer where GSDL is installed, department folders were created. Reports were downloaded from departmental computer to their respective folder in library computer for developing a database. For example, reports generated and available in Wood Coating department computer are downloaded to a folder with the name 'WOOD FINISHES

REPORTS' that was created in the library computer where GSDL was installed. This act was 'source data' for processing the files in GSDL. Input source data is specified as

file://c:\woodfinishinesreports\

Similarly the data downloaded from INTERNET, is copied in the source data of concerned departmental folder.

HOW GSDL WORKS?

The "collector function" is the main facility that helps in creating new collections, modifying or adding to existing ones or deleting collections in GSDL. It has a series of sequences that are self-guided to create a collection. The collector mode has five sequences to build a collection.

- Collection Information – Specifies the name of the collection
- Source Data – Specifies the source data
- Configuration Collection – Has an option to adjust the look of the home page
- Build Collection – Builds the collection (Main process)
- View Collection – View the built collection with report option

When building the collection, GSDL processes different format of source document by seeking a 'plugin' that deals with that particular format. Plugins are specified in the collection configuration file. GSDL generally uses the file name to determine document formats, for example foo.txt is processed as text file, foo.html as HTML and foo.doc as a Word file. Figure 2 depicts the collection processing activity of GSDL.

The construction of all collection is controlled by specifications in a special collection configuration file. Advanced users may use this option to alter the configuration settings.

Building stage is an important stage where it indexes for both browsing and searching options. The processing step takes minutes to

hours depending on size of the collection and speed of the computer. When the collection building task is completed the view collection option activates. After building various collections the home page looks as in Figure 3.

Configuration settings have an option to select the pictures for better look of home page. The software has many features for customisation to suit the requirements.

'Paper to collection guide' details on creating CD-ROM collection from paper. It describes

full details on economics involved in scanning and OCR processes, so as to bring a right format to apply to the GSDL. Collection organiser is another software associated with GSDL that helps in creating and editing the material associated with collection.

ADVANTAGES & DISADVANTAGES OF GSDL

The GSDL software is user friendly, menu driven and has easy operations. It has features to extract the metadata from various file formats such as .TXT, .HTM, .HTML, .DOC, .PDF, .PS, .JPEG, .JPG, .DBF, .TIF, and so on.

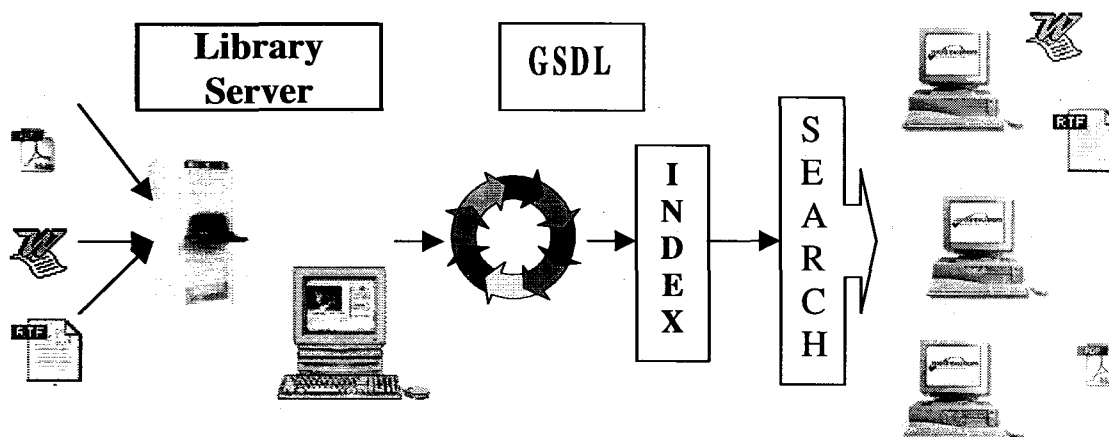


Fig. 2 — Collection and Processing activity of GSDL

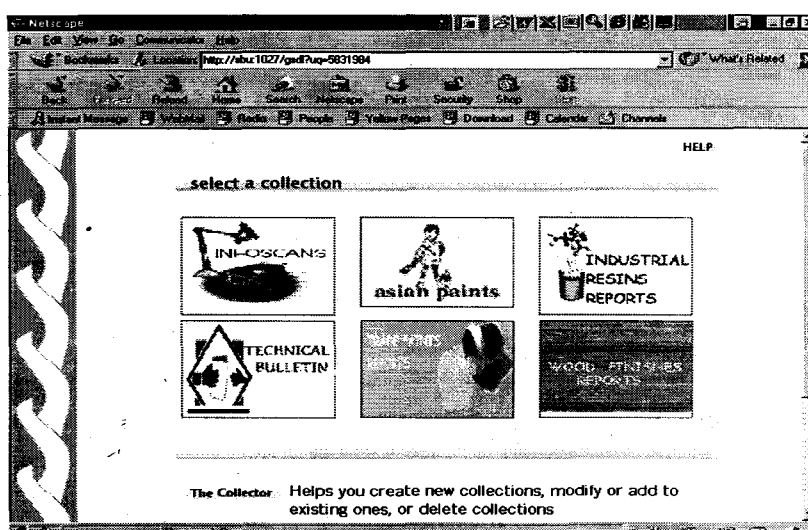


Fig. 3 — Collection Home Page

However, among the disadvantages, addition of file to the database needs reindexing of entire database. Sometimes the software rejects the files while processing and the reason for this is not known. It needs further investigation.

Conversion of existing hard copies into soft version needs a thorough study and analysis in terms of scanner requirements, OCR software, quantity of conversion of documents and manpower training in operations of above equipments. If conversion of hard copies to soft version is a routine, it is better to establish a unit otherwise in case of one time conversion it is advised to outsource the job to professionals. It is recommended to refer GSDL manuals [8,9,10,11] for the entire operations of software and its salient features for better results.

CONCLUSION

Talent, skills, relationships, human networks and institutional knowledge are the great intangible assets of any organization. The concept KM has evolved as a new domain and specific positions like 'Knowledge Managers' are being created. Creating a web of relationships and tapping members' energy, personal time and passion are key aspects of KM to add value to the organisation. It helps in connecting people and builds the network of communities in practice. It facilitates in integrating intellectual value of an organisation and serves as a guide to future employees. "Just in time" is the key aspect of KM generation of future.

In corporate world, transformation of Librarian into Knowledge Manager is clearly taking place.

The present study is only an indicative one for any practitioner for implementing KM in his organisation. There are several softwares and ideas to inculcate the concept of KM in an organisation. Success of KM system depends on the interest of organisations, effective implementation in work practice and each individual's contribution in the community for achieving the company's goal.

REFERENCES

1. <http://www.cio.com/research/knowledge/edit/kmabcs.html>
2. <http://www.destinationkm.com/articles/default.asp?ArticleID=949>
3. SCHULTZE (U) and LEIDNER (D). Studying Knowledge Management in Information Systems Research: Discourses and Theoretical Assumptions. *MIS Quarterly*. 26, 3; 2002; 213-242.
4. MALHOTRA (Y). Tools@work: Deciphering the KM Hype. *The Journal of Quality and Participation*. 21, 4; 1998; 58-60.
5. http://searchdomino.techtarget.com/sDefinition/0,,sid4_gci212449,00.html
6. O'ROURKE (R) and CARROLL (C). Accelerating Knowledge Usage. *European Chemical News*, July 2002; 21-22
7. Ibid. 4.
8. WITTEN (I H) and BODDIE (S). Greenstone Digital Library User's Guide, 2001. Department of Computer Science, University of Waikato; New Zealand.
9. WITTEN (I H) and BODDIE (S). Greenstone Digital Library Installer's Guide, 2001. Department of Computer Science, University of Waikato; New Zealand.
10. BAINBRIDGE (D), McKAY (D) and WITTEN (I H). Greenstone Digital Library 2001. Developers' Guide, Department of Computer Science, University of Waikato; New Zealand.
11. LOOTS (M), CAMARZON (D) and WITTEN (I H). Greenstone Digital Library From Paper to Collection. 2001. Department of Computer Science, University of Waikato; New Zealand.