

KIIT Digital Library: An open hypermedia Application

Dr. Shakeel Ahmed Khoja
Assistant Professor,
KIIT, Karachi
shakeel@kiit.edu.pk

Javed Anjum Sheikh
Librarian,
KIIT, Karachi
anjum@kiit.edu.pk

Abstract

The massive use of Web technologies has spurred a new revolution in information storing and retrieving. It has always been an issue whether to incorporate hyperlinks embedded in a document or to store them separately in a linkbase, known as Open Hypermedia Systems. Research effort has also been concentrated on the development of link services that enable hypermedia functionality to be integrated into the general computing environment and allow linking from all tools on the browser or desktop. KIIT digital library is such an application that focuses mainly on architecture and protocols of Open Hypermedia Systems (OHS), providing online document authoring, browsing, cataloging, searching and updating features.

The WWW needs fundamentally new frameworks and concepts to support new search and indexing functionality. This is because of the frequent use of digital archives and to maintain huge amount of database and documents. These digital materials range from electronic versions of books and journals offered by traditional publishers to manuscripts, photographs, maps, sound recordings and similar materials digitized from libraries' own special collections to new electronic scholarly and scientific databases developed through the collaboration of researchers, computer and information scientists, and librarians.

Metadata in catalogue systems are an indispensable tool to find information and services in networks. Technological advances provide new opportunities to facilitate the process of collecting and maintaining metadata and to facilitate using catalogue systems. The overall objective is how to make best use of catalogue systems. Information systems such as the World Wide Web, Digital Libraries, inventories of satellite images and other repositories contain more data than ever before, are globally distributed, easy to use and, therefore, become accessible to huge, heterogeneous user groups. For KIIT Digital Library, we have used Resource Development Framework (RDF) and Dublin Core (DC) standards to incorporate metadata.

Overall KIIT digital library provides electronic access to information in many different forms. Recent technological advances make the storage and transmission of digital information possible. This project is to design and implement a cataloguing system of the digital library system suitable for storage, indexing, and retrieving information and providing that information across the Internet. The goal is to allow users to quickly search indices to locate segments of interests and view and manipulate these segments on their remote computers.

1.0 Introduction

A digital library is a collection of information that is stored and accessed electronically. The purpose of a digital library is to provide a central location for accessing information. As new information is created and shared among the millions of Internet users at an awe-inspiring rate, it is possible that most of the information for which we are looking exists on the net somewhere. All we require to do is to use an effective searching technique to find them. Unfortunately, the supply of information is increasing far more rapidly than our ability to support effective searching of this huge resource. Therefore, we must first

make fundamental advances in how this information can be captured, stored, searched, filtered, and displayed before it grows out of control.

In 1965, J. C. R. Licklider coined the phrase "library of the future" to refer to his vision of a fully computer-based library [1], and ten years later, F.W. Lancaster [2] wrote of the soon-to-come "paperless library." About the same time Ted Nelson [3] explored Internet and came up with the terms hyperspace and hypertext.. He also analyzed some of the problems to be identified later in his project in some detail named 'Xanadu', but never built an operational system. Many other terms have been coined to refer to the concept of a digitized library, including electronic library, virtual library, library without walls and bionic library [4].

This paper addresses the standards for developing a digital Library. These standards describe the file formats, storage and retrieval standards for digital objects created by or incorporated into the part of the permanent collections. They attempt to balance adherence to industry standards, reproduction quality, access, potential longevity and finally reduction cost. These standards are not intended to address all of the administrative, operational, and technical issues surrounding the creation of digital object collections. Internet access does not require detailed cataloguing. It is important to improve domain-specific indexing but the real challenge is cross-domain searching, which has been tried in KIIT Digital Library (KIIT-DL). Further we have implemented these standards throughout the developing phase of KIIT-DL.

3. KIIT Digital Library: Aims and Scope

The primary aim of KIIT-DL is to provide coherent organization and convenient access to typically large amounts of digital information. The following principles provide working definitions of a digital library from both a conceptual and a practical standpoint:

- KIIT-DL is an integrated set of services for capturing, cataloging, storing, searching, protecting, and retrieving information.
- KIIT-DL services bring order where data floods and information mismanagement have caused much critical information to be incoherent, unavailable, or lost.

- KIIT-DL architecture emphasizes organization, acquisition, preservation, and utilization of information.
- KIIT-DL system can be considered as realization of an architecture in a specific hardware, networking, and software situation, that is specific to any educational organization of Pakistan.
- KIIT-DL heavily relies on metadata standard such as Dublin Core (DC) and Resource Development Framework (RDF), which is discussed in section 4.

3. KIIT-DL Framework

KIIT-DL primary focuses on structured method of storing digital archive, along with searching techniques are provided for quick access of data. This approach is aided with metadata, which provide external classifying and organization of relations for data that can be unstructured and complex in nature. Middleware services such as search, asset protection, and retrieval processes are depended on metadata. Since metadata refers to data, which may be stored in separate hierarchical storage subsystem, integrity of reference must be maintained between metadata and data.

Features such as Content Acquisition (Collection development policies and defining the scope of collections as well as right management) Content Description (Cataloging, metadata, authority control) and Content Access (design, audience, relationship with user via content) are also provided in KIIT-DL. Figure 1 shows the basic features of KIIT-DL.

3.1 Database Model

A client-Server approach is used for the development. At the Server end, Apache Server is used because of its robust and stability. The client will be using a typical WWW browser (Internet Explorer, Netscape, Mosaic etc) and does not require any plug-in or client software to install, whereas in the middle ware, we are using PHP. PHP is open-source server-side HTML-embedded scripting language. Proprietary products in this niche are MS ASP, Cold Fusion, and JSP. PHP is free, easy, embedded, cross-platform, fast, stable, popular and growing scripting language. The database used for the product is MySQL. MySQL started out as a tool to fulfill an internal requirement. MySQL has grown in popularity over the years. It is a fast database and implements a subset of SQL. It is a

standardized way of maintaining the database of any sort, regardless of the underlying methods of saving and retrieving data. Finally it is free to use.

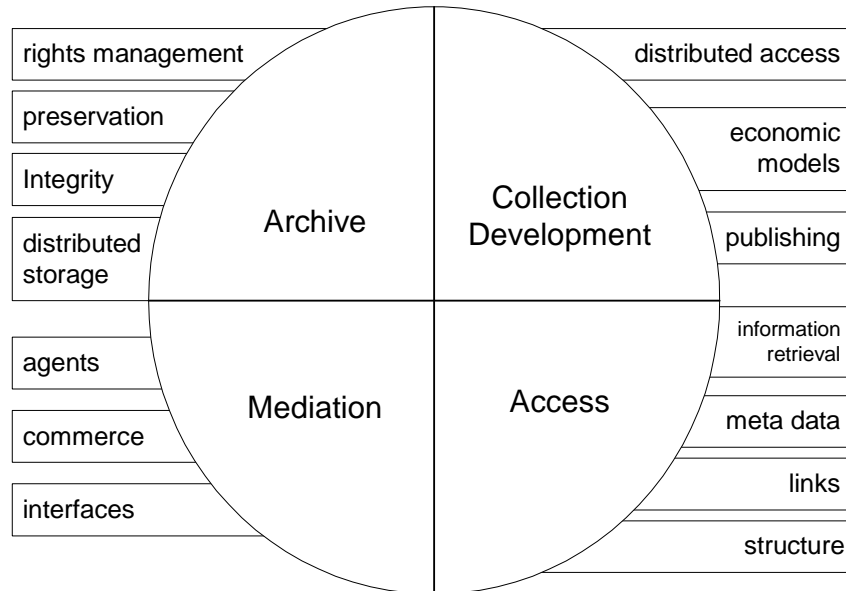


Figure1: Basic Data Access Features of KIIT-DL

3.2 KIIT-DL as Open Hypermedia System (OHS)

An OHS is typically a middleware component in the computing environment offering hypermedia functionality to applications orthogonal to their stage and display functionality. An important matter in OHS is the distinction between structure and content i.e. hypermedia links or structural data (metadata in our case) has to be stored outside the document and are not embedded within the document[6]. This holds true for KIIT-DL, as all the metadata is stored in XML form (shown in table 1) as a wrapper to the document, but not embedded in the document. KIIT-DL also allows different types of documents to be stored in their own formats and later accessed through their own plug-in. This also fulfills another aspect of being an OHS.

Link traversing in KIIT-DL is conducted as:

1. The user clicks on an anchor in the HTML file displayed on the front end of KIIT-DL browser. By definition, the URL for the other end of the link is embedded in the HTML file being displayed.
2. Based on the information of URL, the KIIT-DL sends a 'get' request to the server holding the document.
3. The server sends the file back to KIIT-DL.

4. Based on the type of the file, KIIT-DL browser will do one of the following:
 - a. HTML file: the document contents will be interpreted and the document will be displayed with outgoing links if any. It may now be possible for the user to follow links to other files (forward traversing).
 - b. Other file types: the KIIT-DL file can itself display some files types (.doc, .ppt, .gif, .jpg, .swf, etc.) but these type of files will have no outgoing links. If the browser does not open any type of file, it will start gathering the information for the plug-in and will attempt to download it from WWW.

Like other OHS[7], KIIT-DL allows an open set of applications to participate in the hypermedia service and supports an open set of data model formats.

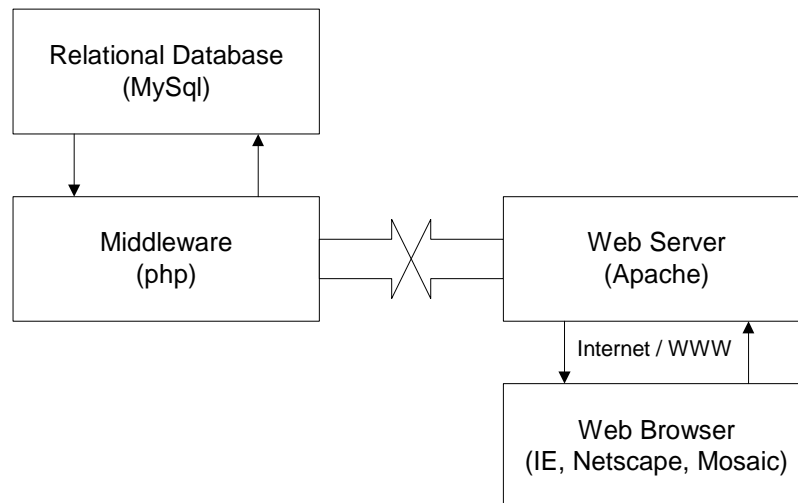


Figure2: KIIT-DL System Architecture

3.4 User Interface

Figure 3.1 shows the main page of KIIT-DL. Here the user can select different options like adding an article, browsing the categories, quick search, and login to the main server. Another text box is provided for the ad hoc search, so that user does not have to login to the system. Figure 3.2 depicts the topic category page for KIIT-DL. Here the topics are categorized as, newspaper clippings, course materials, proceedings, e-books, publications, etc.

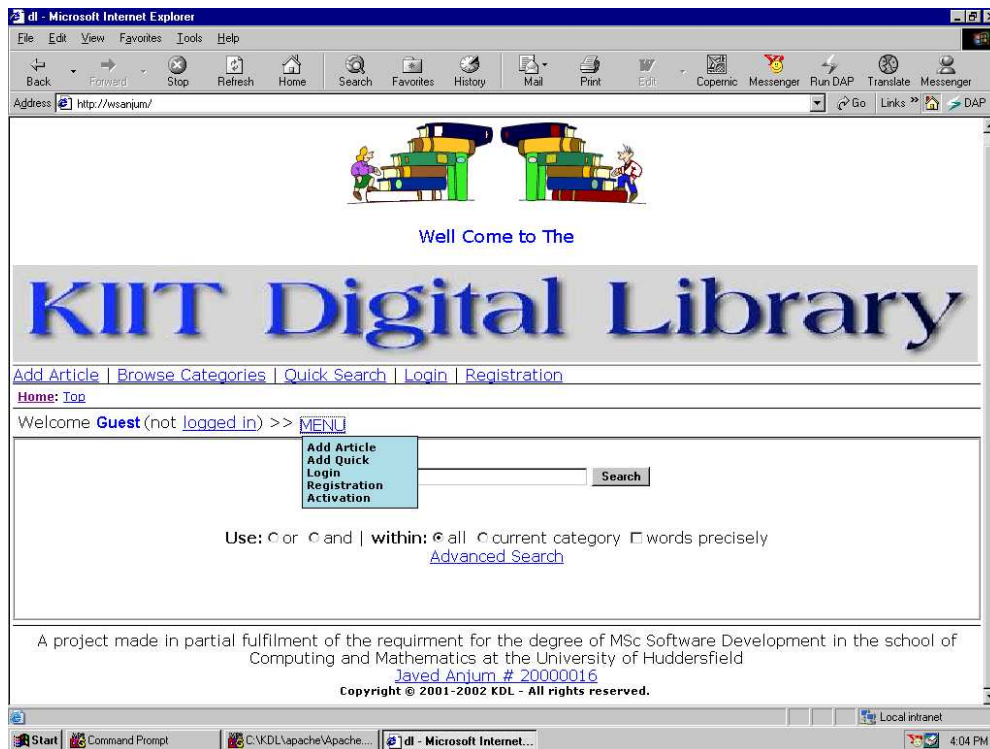


Figure-3.1: Front end of KIIT-DL

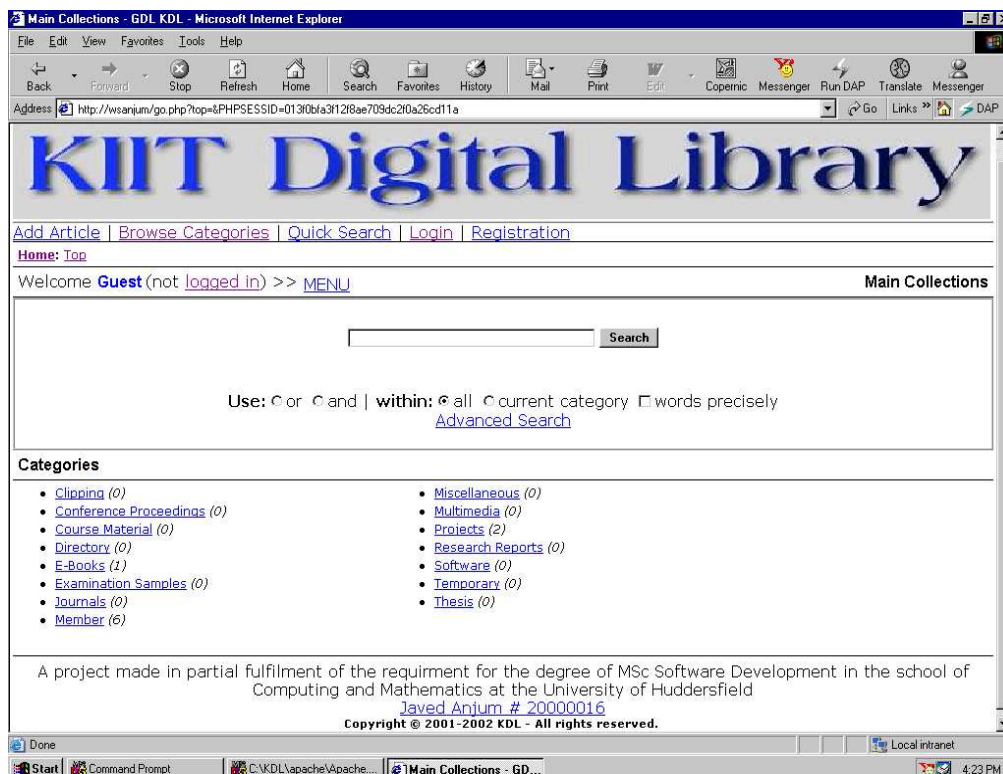


Figure-3.2: Topics Category Page of KIIT-DL

4.0 Incorporation of Metadata Standards in Digital Library

The most common problem with Digital libraries (DL) is that users of DLs usually have to browse through large amounts of data in order to find informative, i.e. relevant, parts of documents. Browsing means -in this context- that users want to scan quickly through data to inspect and compare information they specified only roughly by a previous query. A Digital Library system has to provide support for this kind of information-intensive work. Hence, the access to data should be based on task-related, conceptual, and content-based criteria, especially for time-dependent data like video and audio, whose presentation is very time-consuming. In a browsing application users need efficient ways to access those parts of videos that contribute to this relevance in their given context.

Browsing applications not only have an impact on the required system's data access functionality, but also on the presentation and interaction capabilities. Depending on the users' preferences, some data are more and some less important. The main goal of users in digital library applications is to find information quickly, rather than to actually view documents as in Open Doc applications, since they will not have time to systematically go through and view the entire data. The system has to support efficiently the browsing process by means of automatic selection and structuring of the relevant data.

An important feature of KIIT-DL is the standardization of metadata. Resource Description Framework (RDF) and Dublin Core (DC), designed by World Wide Web Consortium (W3C), standards are used in this regard. RDF is a specification currently under development within the W3C Metadata activity (<http://www.w3.org/Metadata>). RDF is designed to provide an infrastructure to support metadata across many web-based activities. It is the result of a number of metadata communities bringing together their needs to provide a robust and flexible architecture for supporting metadata on the internet and WWW. Its design has been heavily influenced by the Warwick Framework work [5]. In our system, RDF is used to provide a uniform and interoperable means to exchange the metadata between programs and across the web. It also provides a means for publishing both a human-readable and a machine-understandable definition of the property set itself. RDF also allows different application communities to define the metadata property set that best serves the needs of each community. XML (Extended Markup Language) is used as the transfer syntax in order to enforce the other tools and code bases being built around XML. For example, SAMI (Synchronized Accessible Media Interchange),

developed by Microsoft, and SMIL (Synchronized Multimedia Integration Language), developed by W3C (<http://www.w3.org/TR/WD-smil>), are the examples of web-based multimedia presentations encoded in XML. Generally, the goal of RDF is to define a mechanism for describing resources that makes no assumptions about a particular application domain, nor defines (a priori) the semantics of any application domain. The definition of the mechanism should be domain neutral, yet the mechanism should be suitable for describing information about any domain.

Dublin Core (DC) was designed specifically for generating metadata for textual documents. (<http://www.purl.org/DC>). It is a cross-disciplinary international effort to develop mechanisms for the discovery-oriented description of diverse resources in an electronic environment. The Dublin Core Element Set comprises fifteen elements (DCES), which together capture a representation of essential aspects related to the description of resources.

The majority of work on the Dublin Core has addressed the definition of semantics rather than syntax or structure, allowing rapid conceptual development free of the constraints imposed by specific implementation environments. Whilst beneficial in many ways, this has led to a certain lack of clarity at times, especially in relation to the development of 'qualification' mechanisms, which enrich descriptions in the Dublin Core. It has also made interoperable implementation difficult, as individual implementers have typically developed their own internal mechanisms for actually encoding Dublin Core; mechanisms which are not always compatible with those of their potential collaborators elsewhere.

The elements of Dublin core are: Title, Creator, Subject, Description, Publisher, Contributor, Date, Type, Format, Identifier, Source, Language, Relation, Coverage and Rights. A complete description and definition of these elements is available at Dublin Core Web Page. (<http://www.purl.org/dc/documents/dcmes-qualifiers>).

For searching through keywords throughout the document, we have implemented two main strategies for accessing metadata. First is a simultaneous keyword search of all metadata in all repositories, corresponding to full range of unqualified Dublin Core elements. This will result in for of a interwoven grid of the metadata for each repository (holds true for audiovisual media) and then the user is finally provided with the list of best hits. In the second method, we simply search throughout the metadata and the

content itself to perform an exhaustive search (applicable to text documents only) to provide the use a bigger range of selection. The first approach results in a highly structured way to describe stored contents and make them easy to locate and retrieve. Figure 4 shows a snapshot of RDF and DC metadata for a stored document in KIIT-DL. Similarly table 1 shows an XML schema generated for adding a document to KIIT-DL.

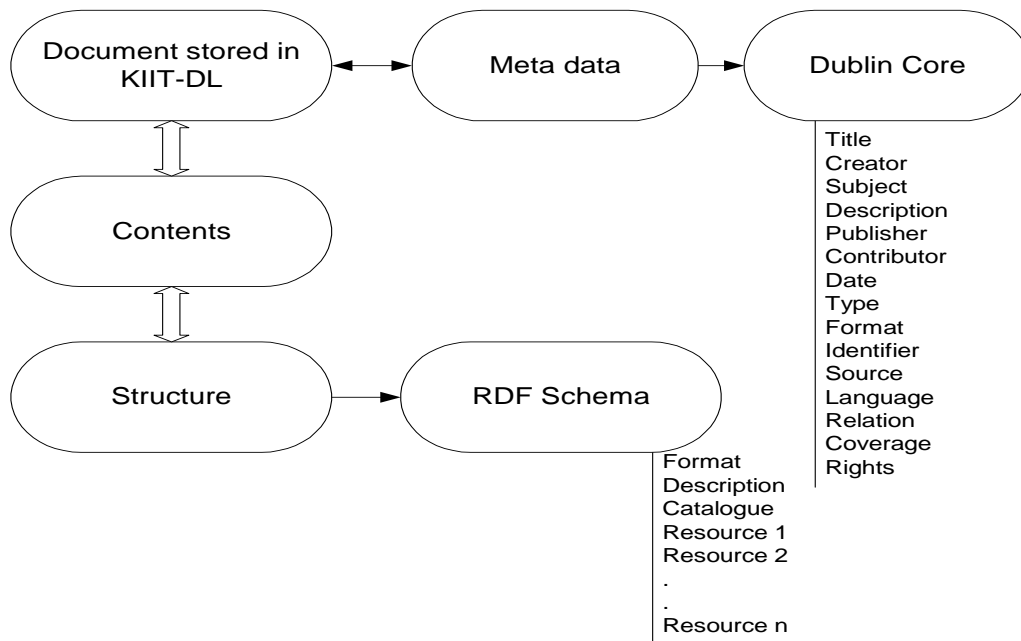


Figure-4: Metadata Model for a document in KIIT DL

5.0 Conclusion

The vision of KIIT-DL is to establish a prestigious digital library, which can host research papers of international repute (plans for establishing research journal are in the pipeline), store online course contents for a university, a linking hub for web services and information, archive original manuscripts and repositories and should be of low cost to the readers, along with integrating metadata and WWW standards. At the moment we are running this digital library successfully for around 400 students and 50 faculty members. KIIT-DL plans to introduce a unique theme-based editorial structure as a means of managing and integrating papers within its broad scope. Each theme will have a theme editor. Themes could evolve to become journals within a journal, but will act cooperatively rather than competitively, and papers are interlinked and cross-indexed for ease of access. KIIT-DL, along with other pioneering electronic-only journals, bases presentation on open, published standards and freely-available software. The aim is to

make it browser-independent, and to restrict the need for browser plug-ins and software downloads to supplementary data. Most important, aims are to explore the new dimensions of electronic writing and publishing: space, through the use of distributed, linked information; and time, to discover how information can evolve to serve today's user.

6.0 References

1. Licklider, J. C. R. Libraries of the Future. Cambridge, Mass.: M.I.T. Press, 1965.
2. Lancaster, F. Wilfrid. Toward paperless information systems. New York: Academic Press, 1978.
3. Nelson, Theodor H. Computer Lib. Chicago: Nelson Press USA, 1974.
4. Drabenstott, Karen. Analytical Review of the Library of the Future. Council on Library Resources; Washington, D.C.
5. Lagoze, C., The Warwick Framework, a container architecture for diverse sets of metadata, D-Lib Magazine, July/August 1996, ISSN: 1082-9873
6. Hodge, Gail. Best practice for digital archiving: an information life cycle approach. D-Lib Magazine, January 2000] .
7. Baldonado, Michelle, Chen-Chuan Chang, Luis Gravano, Andreas Paepcke. Metadata for digital libraries: architecture and design rationale. In: Proceedings of the 2nd ACM International Conference on Digital Libraries, pp. 47-56, 1997

Table 1

```

<?xml version="1.0" ?>
<KDL>
<dc.title>
KIIT digital Library: An Introduction
<about>KDL</about>
<alternative>Library on net</alternative>
<series>Digital</series>
</dc.title>
<dc.creator>
Javed Anjum Sheikh
<bday>1971-04-30</bday>
<orgname>Karachi Institute of Information Technology</orgname>
<email>anjum@kiit.edu.pk</email>
<street>1434/3 S. F. Colony</street>
<locality>Karachi</locality>
<region>Sindh</region>
<country>Pakistan</country>
<pcode>75230</pcode>
</dc.creator>
<dc.publisher>
<id>KDL</id>
</dc.publisher>
<dc.subject>
Computer
<keywords>Digital Library RDF Matadata PHP DC Dublin Core MySQL Apache</keywords>
<classification>Library Science</classification>
<ddc />
</dc.subject>
<dc.description>
This is introduction about KIIT digital Library
<alternative>This is first about KIIT digitalLibrary</alternative>
<notes>word file</notes>
</dc.description>
<dc.contributor>
Zia, Director
<role>Digitalize Person</role>
</dc.contributor>
<dc.contributor>
Akbar, Manager
<role>Editor</role>
</dc.contributor>
<dc.contributor>
Usman, Asstt.
<role>Translator</role>
</dc.contributor>
<dc.contributor>
Farzan, System Administrator
<role>Programmer</role>
</dc.contributor>
<dc.contributor>
anjum@kiit.edu.pk
<role>Content Editor</role>
</dc.contributor>
<dc.contributor>
<modifiedby>anjum@kiit.edu.pk</modifiedby>
</dc.contributor>
<dc.date>
2002-02-28
<issued>2002-02-28</issued>
<modified>2002-02-28-04-34</modified>
</dc.date>
<dc.type>prj</dc.type>
<dc.format />
<dc.identifier>kdl-kdl-prj-2002-javed-9-kdl</dc.identifier>
<dc.identifier>
<hierarchy>Member/anjum@kiit.edu.pk</hierarchy>
</dc.identifier>
<dc.source />
<dc.language>English</dc.language>
<dc.relation />
<dc.relation>
<count>1</count>
</dc.relation>
<dc.relation>
<no>1</no>
<haspart>kdl-kdl-prj-2002-javed-9-kdl-rightup.doc</haspart>
<hasformat>application/msword</hasformat>
<hassize>59904</hassize>
<hasuri>/download.php?f=/disk1/1/kdl-kdl-prj-2002-javed-9-
kdl-rightup.doc</hasuri>
<hasfilename>rightup.doc</hasfilename>
</dc.relation>
<dc.coverage />
</KDL>

```

//continued in col 2