Developing databases on and for ODL research

Prof. Zhu Zhiting PhD
Vice Dean, College of Online Education and Director, Educational Information Center
East China Normal University, Shanghai, China (200062)
E-mail: ztzhu@dec.ecnu.edu.cn; ztzhu@hotmail.com

[Abstract] This paper discusses emerging interest among ODL researchers in creating an integrated information environment, so that ODL research and learning activities can be facilitated effectively through the utilization of the vast amount of Internet information resources. Starting with an analytical review of problems with the utilization of distributed resources, this paper first investigates technological solutions applicable to the identified problems, then follows with a proposition for a knowledgeable system framework that can serve as a conceptual model for the development of ODL research-oriented information environments. A number of critical elements composing the model, such as partnership, knowledge management, intelligent agents, personalized information services and standards application, are further described.

[Keywords] ODL, Database, Internet information resources, research oriented information system, system framework, knowledge management

Introduction: ODL Researches and the Utilization of Internet Information Resources

Open & Distance Learning (ODL), which has contributed greatly to expansion of the scope of education in temporal and spatial dimensions, is drawing considerable amount of research efforts in many countries. It seems that information resources regarding ODL, which present in forms such as websites, e-journals, databases, software, newsletters, digital video/audio recordings and text documents, are proliferating rapidly in cyberspace. While traditional library services seem to meet the needs of face-to-face teaching and learning and its associated research, information on the Internet, although broader in scope, may not suffice the need of learning and researching online. Firstly, as the information resources are spread widely over the Internet and exist in huge amounts, most search engines can only cover about 15% of

1 This author thanks Mr. Hu Xiaoyong for his assistance in drafting the first manuscript of this paper.
web pages. Secondly, the results retrieved are not tailored to the specific requirements of different kinds of users including researchers, teachers and learners. Information users have to select objects from a long list of items. These problems imply an urgent need for the development of innovative information systems that can integrate distributed resources and provide domain-specific and personalized information services.

Given the topic by the organizers of the CRIDALA 2002 conference, I have struggled with this paper for some months. Based on a review of existing efforts regarding the integration and applications of distributed resources, this paper will firstly look into technological approaches for dealing with the problems encountered in utilizing distributed online resources. Further, this paper will propose a framework for developing a knowledgeable information system facilitating ODL research. In my understanding, the concept of *databases* here is best understood as a domain-specific (ODL in the context of this paper) and research-oriented information environment, not simple collections of data in the traditional sense.

**General Solutions to the problems of Utilizing Internet Information Resources : A technological analysis**

It is necessary to identify the main sources causing difficulties in utilizing Internet information resources. The first source of difficulty is the distributivity of Internet information resources, because distributed resources cause low-level of structurality and thus users experience problems in retrieving and utilizing them. In a global view, Internet information resources are rather ill-structured and/or ill-ordered, though they might be well-organized locally within a specific database.

The second source of difficulty is the heterogeneity of Internet resources, because heterogeneous resources cause interoperability problems. According to Sheth[1], interoperability problems could appear at four different levels: the system level, syntactic level, structural level and semantic level. The system level includes incompatible hardware and operating systems; the syntactic level refers to different languages and data representations; the structural level includes different data models and the semantic level refers to the meaning of terms using in the interchange.

These rational analyses lead to a classification of Internet information resources in the light of two dimensions: heterogeneity and structurality. As shown by Fig.1, in the dimension of heterogeneity, there are differences between homogeneous resources and heterogeneous resources, while in the dimension of structurality, there are differences between well-structured resources and ill-structured resources.

- Homogeneous & well-structured (Area I): A collection of resources which have consistent formats and are well-organized. These resources are frequently organized into a specific database and operated in similar
environments (nearly same type of platform, database system, format, etc.).

- Homogeneous & ill-structured (Area II): Referring to resources that exist fragmentarily in the cyberspace but are operated in similar environments (nearly the same type of platform, database system, format, etc.).

- Heterogeneous & well-structured (Area III): Referring to resources whose structure of content is integral and stable, but they are operated in different environments (different types of platforms, database systems, formats, etc.).

- Heterogeneous & ill-structured (Area IV): It means various resources not only whose structure of content is discrete and bitty, but also they operate in different environments (different platforms, database systems, formats, etc.).

![Figure 1: Technological approaches to the integration of distributed resources](image)

It would be more desirable for all resources to be well-structured and stored in a single well-organized database. However, this situation can only happen in the settings of LANs or intranets. Due to the diversity and distributivity of online resources, never can we expect all resources to be homogeneous and well-structured.

There have been some technological approaches for harnessing problems with the integration of distributed resources. As shown in Fig.1, the first kind of technology is relevant to structuralization, which contributes to turning fragmentary and chaotic data sets to ordered and structured representations. If not possible to structurize the huge body of various data sets, a metadata-based mechanism can at least be created, e.g., a centralized catalogue or indexing system. Another option is to create a gateway or information broker that queries a number of databases and returns all the retrieved data to the user. However, this approach still leaves the user with the decision of
which search engines to use, and interpreting the results for each engine separately.

The second kind of technology is relevant to interoperation, which contributes to resolve heterogeneous problems of Internet information resources. Many technologies have been developed to tackle heterogeneity at system level, syntactic level and structural level, which have been addressed using technologies such as CORBA, DCOM and various middleware products. Recently XML has gained acceptance as a way of providing a common syntax for exchanging heterogeneous information. For the heterogeneity at semantic level, a recent solution is the use of ontology to formally specify the meaning of the terminology of each system and to define a translation between each system terminologies and an intermediate terminology [2].

Developing databases on and for ODL research: A framework of knowledge management

It is not sufficient to build a useful information system facilitating ODL research if only relying on the afore-mentioned interoperation and structuralization technologies. Further questions arise, such as:

- How do we effectively and efficiently integrate global education networks and gateways into research about teaching and learning in ODL?
- What are the best ways to integrate online research designs with databases, data collections, and data analysis?
- Where and who should claim ownership of the emerging fashion of regionally positioned databases in strategic places in the world?
- What new protocols would need to be developed for coping with the manipulation of huge data sets that are emerging through the virtual space of teaching and learning?
- What are suitable strategies for the sustainable development of a research-oriented information system?

Based on substantial investigations into the fields regarding the integration of distributed resources and the development of research-oriented information systems, this author is proposing a knowledgeable framework for building an information environment facilitating ODL research, in which partial answers to afore-mentioned questions are involved. As shown in Fig.2, the framework integrates five important elements: partnership knowledge management, intelligent agents, personalized information services, and standards application.
Fig.2 A knowledgeable framework of ODL research information system

**Partnership**

It is assumed that the developers of this proposed system are ODL researchers supported by their affiliations that usually possess considerable amounts of information resources regarding ODL. They are now willing to work in partnership to create an integrated information system facilitating ODL research and also favoring other users in the field of ODL. Particular developers might play multiple roles, e.g., a researcher who contributes to proliferating the system’s information repertoire, or a user who receives information services provided by the system. It is necessary to design a policy specifying users’ rights for future accessing/consuming information from the system. A basic principle is the more one contributes the more services one is privileged to receive, which can be called *utilization protocol*. A *user profiling*
mechanism will record each user’s information contributions and consuming preferences so that the utilization protocol and personalized services can be implemented. Therefore, the partnership seems to be a realistic answer to the ownership problem caused by sharing distributed resources.

Knowledge Management

Knowledge Management is understood as the process of creating, archiving, and sharing of valued information, expertise, and insight within and across communities of people and organizations with similar interests and needs [3].

Nonaka [4] defined two types of knowledge: (1) tacit knowledge, which is subjective and experience based that can’t be easily expressed in words, sentences, numbers or formulas, often because it is context specific; and (2) explicit knowledge, which is objective and rational knowledge that can be expressed in words or sentences. He has suggested that knowledge is created through four different modes:

- **Socialization**, which involves conversion from tacit knowledge to tacit knowledge,
- **Externalization**, which involves conversion from tacit knowledge to explicit knowledge,
- **Combination**, which involves conversion from explicit knowledge to explicit knowledge, and
- **Internalization**, which involves conversion from explicit knowledge to tacit knowledge.

Milton *et al* [5] suggested that various structures of knowledge technology could be developed to support five key knowledge management activities: personalization, creation/innovation, codification, discovery and capture/monitor.

In our framework, knowledge management is taken as a suitable strategy for the sustainable development of research-oriented information systems. Multiple structures of knowledge technology are involved in support of knowledge transfers between different modes. Knowledge externalization is facilitated by the research registration mechanism that serves to capture researchers’ knowledge generated in their research actions and processes, which can be expressed in forms of research proposals, progressive reports, expert referees, and so on.

Knowledge combination is fulfilled by an ontology agent in association with a metadata agent. The ontology is an explicit and formal specification of a shared conceptualization of ODL-related domain and offers information as contexts to a metadata agent. The metadata agent captures context-relevant content objects from the repertoire of resources (including those registered research materials) and extracts metadata from the content objects captured as inputs to the metadata repertoire. Meanwhile, the ontology agent extracts key terminologies and relationships from the
metadata repertoire as inputs to the *ODL-specific ontology*.

On the other side, knowledge internalization is facilitated by the personalized information retrieving services, which are provided by a *search agent* in association with the user profiling, ontology, and metadata-based indexing mechanisms.

Furthermore, the processes of knowledge *socialization* taking place among ODL researchers and learning users can be facilitated by various type of CMC tools such as conferencing, e-mail, e-presentation, and so on, which are not overtly depicted in the framework for simplification reasons.

**Intelligent Agents**

New approaches have emerged for the development of information systems that use various types of intelligent agents as their characteristics, such as autonomy, adaptivity, collaboration, and mobility. For instance, Isaias[6] proposed an architecture of digital libraries and electronic publication that is composed of eight different types of agents.

In our framework, the employment of agents presents an efficient and effective approach for dealing with massive resources distributed over the Internet. These agents would undertake many enduring and onerous tasks such as data capturing/mining, information retrieving/filtering, and knowledge extracting/updating. Unexhaustedly, only three main types of agents are presented in the framework:

- **Ontology agent**, which will extract ODL-oriented terminologies, create their conceptual relationships and update the ODL-specific ontology.
- **Metadata agent**, which will capture ODL-related information from distributed resources with the contexts described by the ontology and extracts metadata from the captured information and those registered research documents.
- **Search agent**, which executes a series of retrievals for meeting information demands from users, with domain constraints described by the ontology and in accordance with demands plus preferences of users as the premises for retrieving operations onto the repertoire of metadata-based indexing.

**Personalized information services**

Personalized information services represent recent efforts to disseminate information to users’ personal needs more effectively. A typical personalized information system keeps a profile for each user so that only the information most relevant to a user is identified and presented. Jayawardana[7] suggested two schemes involved in a personalized information environment: Material personalization corresponds to facilities for learners to use library materials according to their
individual requirements such as active consuming and information gathering. Collection personalization, on the other hand, captures the user’s learning context and interest from the material personalization in order to provide a personalized view of the organization of the digital library. Collection personalization, then, includes personalized filtering and personalized retrieving. In our framework, personalized information services are to be implemented by the search agent in association with the user profiling mechanism. However, we still need further studies on the modeling of profiles for different types of users.

**Standards application**

The application of standards is becoming a common practice for the development of internet-based information environments, because they contribute greatly to the interoperability of application systems and the sharability of Internet information resources [8]. In the proposed framework, a number of standards can be applied, such as XML (Extended Markup Language), UDDI (Universal Description, Discovery and Integration), RDF (Resource Description Framework), OPS (Open Profiling Standard), DOM (Document Object Model).

**Endings**

In this paper, a framework for an Internet-based information system facilitating ODL research has been proposed and elements consisting of the system have been introduced. None of these elements are new if they are viewed separately. However, a systematic integration of these elements will make a difference, because this will help to build a highly valuable information environment for ODL researchers as well as learners. We believe that a knowledge management perspective, a knowledge technology approach and a knowledgeable framework are especially valuable for the development of this kind of environment.

It is noticed that the proposed framework doesn’t present an implementable model for the development of an ODL research oriented information system but is useful for eliciting innovative ideas from enthusiastic colleagues.

**References:**


