

Digital library research: current developments and trends

Ali Shiri

The Authors

Ali Shiri, *Senior Researcher at the Centre for Digital Library Research (CDLR), University of Strathclyde, Glasgow, UK*

Abstract

This column gives an overview of current trends in digital library research under the following headings: digital library architecture, systems, tools and technologies; digital content and collections; metadata; interoperability; standards; knowledge organisation systems; users and usability; legal, organisational, economic, and social issues in digital libraries.

The emergence of digital libraries in the 1990s and their overwhelming growth in recent years has opened up new horizons across a broad array of issues related to design, implementation, development and evaluation of digital libraries. Digital libraries have been defined in a variety of ways. For instance, the Digital Library Federation (2002) defines them as:

... organizations that provide the resources, including the specialised staff, to select, structure, offer intellectual access to, interpret, distribute, preserve the integrity of, and ensure the persistence over time of collections of digital works so that they are readily available for use by a defined community or set of communities.

This definition involves three key components, which constitute the theoretical framework underlying digital libraries, namely:

- (1) people;
- (2) information resources; and
- (3) technology.

An investigation into research areas that have recently been explored or found challenging throws up issues in all three areas.

This column starts our examination of the world of digital libraries by examining the major trends and challenges that researchers of digital libraries encountered in three digital library conferences held in the year 2002:

- (1) Joint Conference on Digital Libraries;
- (2) 6th European Conference on Research and Advanced Technology for Digital Libraries; and
- (3) 5th International Conference on Asian Digital Libraries.

These conferences are not only the main international digital library events during the period, they are also distinguished by their coverage of digital library research and development at a global level.

To provide an integrated and coherent view of research trends, the issues discussed in these conferences should be summarised in eight major categories:

- (1) architecture, systems, tools and technologies;
- (2) digital content and collections;
- (3) metadata;
- (4) interoperability;
- (5) standards;
- (6) knowledge organisation systems;
- (7) users and usability; and
- (8) legal, organisational, economic, and social issues.

Architecture, systems, tools and technologies

Within this category lie all technical, infrastructural and algorithmic and system-related components of digital libraries. Some of the key issues here are:

- open networked architectures for new information environments;
- novel search and retrieval techniques such as federated search using data fusion[1], mediator architecture[2] and integrating links and ranking;
- audio-visual and multimedia information retrieval systems;
- content management systems;
- intelligent systems for indexing, abstracting and information filtering;
- harvesting and interoperability technologies; and
- collaborative, visual, 2D and 3D interfaces.

Digital content and collections

This category refers to individual digital objects and to collections of objects in repositories encompassing a variety of materials in different digital formats.

One major challenge with regard to metadata is the diversity of digital information formats

In this category there are challenges associated with digital content: for instance conversion of printed materials into digital format and the creation of digital-only materials for the purpose of a particular digital library. In addition, a host of other issues related to digital content and collections have been discussed. These include:

- collection development strategies, policies and management;
- identifying collections of information which are not accessible or usable because of technical barriers;
- formulating strategies for sustainable and scalable collections;
- encouraging the development of new collections;
- the creation of digital objects and electronic publishing;
- the creation of new genres of digital objects; and
- issues related to digital preservation and Web archiving.

Metadata

Digital collections require well-structured metadata schemes to describe digital objects and content at various levels of granularity. Structural and descriptive metadata[3] are two general classes of metadata of particular relevance. One major challenge with regard to metadata is the diversity of digital information formats and the ways in which they should be described in different collections with different target audiences and uses.

Issues for metadata researchers include:

- human and algorithmic approaches to metadata provision;
- choosing from a wider range of metadata formats;
- applying metadata standards across digital collections;
- metadata harvesting;
- developing metadata extensions for pedagogical purposes; and
- mappings between different metadata formats.

Interoperability

Interoperability is one of the most heavily discussed issues in digital library research. The requirement for interoperability derives from the fact that various digital libraries with different architectures, metadata formats and underlying technologies wish to effectively interact, something they can do through applying a range of common protocols and standards.

The Open Archives Initiative (OAI) protocol (OAI, 2002) is the most widely discussed and investigated standard for cross-repository interoperability. It allows distributed digital libraries to expose their metadata to a wider range of search and retrieval services and also to extract metadata from Web databases. Z39.50 has also been mentioned as another interoperability protocol for online catalogues and other types of information retrieval systems on the Web.

Standards

Standards within the context of digital library research encompass all protocols and conventions that have been set for digital library architectures, collections, metadata formats, interoperability and so forth.

Some types of standards which have been the focus of research include:

- digital collection development standards;
- archiving and preservation standards;
- metadata formats (e.g. Dublin Core, MARC, IMS);
- cataloguing content and indexing standards;
- electronic publishing standards for books, journals and other media, OAI and Z39.50.

Knowledge organisation systems

This category refers to a range of tools used for organisation, classification and retrieval of knowledge in a general sense. Digital library researchers operating in different contexts have investigated the potential of these tools for different purposes.

Some of the applications are:

- use of thesauri and classification systems for cross-browsing and cross-searching across various digital collections;

- creation of ontologies using existing thesauri;
- classification systems and specialised controlled vocabularies to provide a general knowledge-representation facility for digital collections with a diverse range of materials; and
- use of taxonomies to provide unified and organised access to different digital repositories through describing different layers of the digital collections.

One of the challenges is the way in which these tools can interact with each other. To this end research is under way to investigate issues surrounding mappings and interoperability among various knowledge organisation systems (HILT, 2002).

Users and usability

In order to develop usable digital libraries and to improve system design, researchers have addressed user behaviour and user requirements in different contexts including academic environments, schools, government departments and business.

The following areas have been the focus of a number of studies:

- empirical studies of users interacting with digital libraries;
- usability, accessibility and user acceptance of digital libraries;
- user-centred support for learning, teaching and research through the convergence of virtual learning environments and digital libraries;
- human-computer interaction; and
- evaluation of the behaviour of diverse user communities based on their knowledge base, age level, and particular needs.

One of the major challenges in user studies is associated with the methodology and data gathering techniques. Researchers have tried to use a combination of tools and techniques to collect useful data for user evaluation.

Legal, organisational, economic, and social issues

Rights management, intellectual property and copyright issues are all legal aspects of digital libraries. Social issues in relation to digital libraries centre on the ways in which people view digital libraries and their usefulness; and the extent to which they are integrated into people's lives and social activities. There are also economic issues such as commerce, shopping, marketing and business competition, all of which form part of the digital library research discussion. This area touches on topics such as:

- intellectual property in the complex global market;
- legal issues associated with access, licensing, copying and dissemination of digital materials, economic, business and pricing models and strategies; and
- sustainability and survivability, new business models and marketing strategies.

Although these are the main research categories, the list is not exhaustive. Other issues include:

- evaluation issues outwith the users and usability category, reference and question-answering services; and
- the development of different types of digital libraries.

Evaluation is a critical issue in digital library research. While user-oriented evaluation can be discussed in the users and usability category, evaluation also applies to digital library systems, their performance, the underlying technology and the information retrieval techniques utilised.

Research in digital libraries will span a broad range of subjects, disciplines, contexts and communities in the coming years

There are also a number of projects concerned with the design, development and evaluation of various types of digital libraries. For instance:

- digital libraries addressing different target audience such as children or undergraduate students;
- digital libraries addressing geographical locations, i.e. national digital libraries, rural digital libraries and state digital libraries;
- digital libraries addressing a particular subject area, such as computer science, medicine, mathematics and science; and
- digital libraries targeting a particular type of content, e.g. theses and dissertations, music digital libraries and video digital libraries.

Taken overall, these trends and challenges demonstrate the depth and breadth of research and development in digital libraries. The scope and scale of research calls for the participation of people from a wide range of disciplines and communities, including librarians, information scientists, archivists and computer scientists, as well as people working in the areas of technology, medicine, social sciences and humanities.

Based on the outline above, I have listed below a number of broad research areas with a pragmatic view to the current and future trends in digital library research:

- distributed digital libraries and learning environments (convergence of virtual learning environments and digital libraries);
- new digital publishing and preservation environments;
- building digital subject libraries;
- development of digital libraries for particular types of media;
- federation of digital library systems and the use of advanced search and retrieval features e.g. term weighting, query modification, and relevance feedback;
- utilisation of knowledge organisation systems such as ontologies, thesauri, classification systems and taxonomies;
- development of digital libraries for children, undergraduate students and researchers;
- retrospective digitisation of historical collections and cultural heritage preservation;
- inter-institutional collaboration and international cooperation; and
- user-centred evaluation of digital libraries.

To sum up, research in digital libraries will span a broad range of subjects, disciplines, contexts and communities in the coming years. This development reflects the fact that digital libraries will function as institutional entities with a key role to play in our social, cultural and educational activities. Collaborative digital work environments, social digital navigation and new forms of digital environments for academic, educational, publishing and recreational purposes can be viewed as the source for future developments in digital library research.

Notes

- 1 Data fusion is the seamless integration of data from disparate sources.

2 A mediator is a domain-specific tool to support uniform access to multiple heterogeneous information sources and to abstract and combine data from different but related databases to gain new information.

3 Structural metadata is metadata that describes the types, versions, relationships and other characteristics of digital materials while descriptive metadata provides intellectual access to a work such as MARC cataloguing records.