

Digital Libraries: An Overview

Edward A. Fox

Virginia Tech

(adapted from Ch. 11 for IR book,
AW, 1999, ed. Baeza-Yates, Berthier)

Benefits

- Ease of use
- Effectiveness
- “The benefits of digital libraries will not be appreciated unless they are easy to use effectively.” - IITA Workshop report

Introduction

- References: Lesk, D-Lib, ACM DL'9x, ...
- Vision: Wells, Bush, Licklider - challenge
- 4S to provide integrative framework
 - Streams: text, multimedia, commun. channel
 - Spaces: 1D, 3D, vector, thesaurus, GIS
 - Structures: DB, grammars, markup, web
 - Scenarios: stories, reqs, specs, fns, transforms
 - Network traffic: streams with structure carrying a scenario

Definitions

- Library ++
- Distributed information system + organization + effective interface
- User community, collection, services
- Digital objects, repositories, IPR management, handles, indexes, federated search, hyperbase, annotation

Architectural Issues

- Internet middleware
- Independent system / part of federation
- Decompositions vary
 - search engine, browser, DBMS, MM support
 - repository, handle server, client
 - information resources + mediators, bus or agent collection, client with workspace/environment
- Metrics: e.g., for federated search

Document Models, Representations, and Accesses

- Doc = stream + structure + use-scenario
- Multilingual: content, summary, metadata
- Multimedia: structure, quality (oS), search
- Structured: MARC, SGML, by user: MVD
- Distributed collection: Kleisli, CIMI, Z39.50
- Federated search: collecting, picking site(s), parallel search / fall-back, fusing results
- Access: IPR, payment, security, scenarios

Prototype, Projects, and Interfaces

- CODER, MARIAN, ENVISION; CORE
- DLI
 - Document collections: SGML, spatial, video
 - Search: text, x-vocab, image, video
 - Interfaces: DELIVER, DLITE, PAD++
- International: ERCIM, eLib, NZ, AU, SG, ...
- Usability: user-centered, evolving, Allerton, Xerox/visualization, social context/concerns

Standards

- Protocols/federation
 - Z39.50, CIMI
 - Dienst, NCSTRL
- Metadata
 - TEI: inline, detailed (structure in stream)
 - MARC: two-level, fine-grained
 - Dublin Core: high-level, 15 elements
 - RDF: describing resources/collections, annotation

Future Challenges

- Role of author
 - DC vs. TEI
 - Training, quality, amount of effort, QC
- Worldwide DLs
 - Performance
 - Multilingual
 - Identifying specialty areas in large repositories vs. large numbers of focused collections
- Scaling: personal to global, sustainable