Future Library Services

ARD Prasad
Documentation Research and Training Centre
Indian Statistical Institute
Bangalore
ard at drtc.isibang.ac.in
15th February 2010, LISA VI Conference

Common nightmares

- What would be the future of libraries?
- Will there be libraries in future?
- Less and less clientèle are visiting the library?

My Guess

(many might have had better guesses)

- As long as mankind exists, people do require information.
- As there will be more information glut/overload, people would require information services
- End-users would prefer information intermediaries for search, retrieval and **filtering**
- The **Form** of library & its services will change, **essence** of librarianship may not change (Form Vs. Content)

Open Access

- Open Access to Information is already a phenomenon

- Soon, more subscription based journals would yield to pressures of Open Access Movement
- Web 2.0 facilitates individuals to have space on the web – more information & more noise

Digital Preservation

- Digital Preservation would be a major challenge
- Operating systems, application software, file formats would be constantly changing
- Tentative Solution: Using Open standards
- Even then we have to migrate the content across versions, software, formats.

The Web

- A major medium for
  - Information Generation
  - Storage
  - Retrieval
  - A host of information services

Pitfalls of Web Search Engines

- Too much noise in the search results
- Limited search features:
  - no truncation, but wild cards allowed though not...
    - Ex: "friends like * enemies" results in “with friends like you who needs enemies” - proximity)
  - does not support Boolean nesting
- May not index full-text of the web page or pdf, odt etc.
- May search for plural/singular, synonyms, and grammatical variants without telling you
- Not case-sensitive

Cataloguing is Missing in web documents

- Limiting search (uses tricks, but not metadata)
  - By language (though not all)
  - By country
  - By format of the document
  - By host/domain
- NOTE: there is no field level search like --
  - by author
  - range searches by date etc.

Problem of Search Engines

- It is not that search engines are ineffective
- The ineffectiveness lies in web pages they index

Context Sensitive Search

- Can we do Context sensitive search?
- LIS has many models
  - PRECIS
  - POPSI etc.

LIS Failures

- Though we developed context sensitive retrieval models, We could not develop computational models for context sensitive retrieval models
- Inferencing is missing even in context sensitive retrieval models
- Did we overemphasize on Statistical analysis in Information Retrieval rather than linguistic analysis?

Context Free Searches

- Both Google & Bibliographic databases lack context-sensitive search
- Both Searches heavily depend on keywords
  - Uniterms, phrases
  - Boolean operators (post-coordination)
- Resulting in Context-free search
- So called Full-text search have their pitfalls
- Are we overemphasising on Recall?

Can we?

- Can we get precise search results for queries like
  - When did the Kalinga war take place?
  - What are the trains between Mumbai and Delhi between 8 am to 6 pm?
  - Who are the co-authors of Ranganathan?
  - List the names of persons from IUCAA, who published more than 50 articles.

Information on the Web

- Information on the web is
  - for human understanding
  - You do not get precise information
- Information on the Web is
  - NOT for machine consumption
  - NOT without Noise
  - NOT specific but a gross web page

Semantic Web: one solution?

- Use RDF: to describe web resource
- Use Dublin Core in RDF for documents
- Extend metadata to other object/entities (beyond Dublin Core)
- Use HTML/XHTML only for display of information
- Use ontologies to find relationship among concepts and consequently documents/objects
- Use Inference Engines for deducing information

Solution is Interdisciplinary

- Web Technology +
- Library and Information Science +
- Artificial Intelligence

How to add Semantics to WEB

Some Tools

- RDF (Resource Description Framework) for description
- OWL (Web Ontology Language) /SKOS (Simple Knowledge Organization System) for thesaurus/classification
- Dublin Core for Description of Web Documents
- Many metadata schema on lines of Authority files ex: personal, corporate authors etc. and many more

RDF

- RDF is all about making statements about web resources
- Expressed in **triples** for logical deductions
- A triple contains **subject, predicate** and **object**

Triples – common example

- Siddhartha was the founder of Buddhism
- Siddharth is the **subject**
- was-Found-of is the **predicate**
- Buddhism is the **object**
- One can answer queries like -
  - Who was the founder of Buddhism?
  - What was founded by Siddhartha?
  - Was Buddhism founded by Siddhartha?

Bibliographic data in Triples

- “Elements of classification”, “URI is” “http://hdl.handle.net/1849/456”
- “Ranganathan”, “wrote”, “http://hdl.handle.net/1849/456”
- “http://hdl.handle.net/1849/456”, “was published by”, “Ranganathan Endowment”
- ...
- ...

Metadata

- Can be classified as
  - Bibliographic Metadata
  - Non-Bibliographic Metadata
    - Metadata of Individuals (like Name Authority files)
    - Metadata of Institutes (like Corporate authority files

Metadata

- Metadata can be prepared for anything on the web, Metadata about
  - Services
  - Products
  - Software
  - Processes
  - Events

Caveat

- XML based systems are rich in expressiveness, BUT…
- Xml based systems (like owl or skos) do not scale. They can be used as data exchange formats.

Present Semantic Web

- Presently machines can not automatically generate RDFs
- RDFs , Metadata etc. are to be prepared manually
- SemWeb complements the present web, may not replace it

Present Technologies

- Library Management Software
- Digital Repositories
- Harvesters
- RSS and Atom feeds
- Z39.50, SRU/SRW (India lags behind)
- Federated Search Engines
- Consortia Based Software (ex. CUFTS, GODOT)

Other Technologies for LIS

- Data/ Text Mining
- Knowledge Organization
- Personalization
  - Folksanomies
- Visualization
- Adding Comments
- etc.

Services by DRTC

1) Librarians' Digital Library (LDL): https://drtc.isibang.ac.in

2) Search Digital Libraries (SDL): http://drtc.isibang.ac.in/sdl

3) LIS Journal Content Pages: http://drtc.isibang.ac.in/news

4) Federated Search Engine: http://drtc.isibang.ac.in/cgi-bin/dbwiz2.pl

Demos by DRTC

- Koha (Library Management Software)
- Z39.50 server
- CUFTS (Consortia based ERM - Electronic Research Management System)
- GODOT (Link Resolver, not an article level resolver like SFX)

Open Mantra

- Open Access to Information
- Using Open Source Software
- Open Standards
- Open Learning
- Build open access system using open source software and open standards

Thank you

ard@drtc.isibang.ac.in