Research in digital libraries

As done from human and technological perspectives

Tefko Saracevic, Ph.D.

ToC

- Research orientations & agendas
- Human-oriented digital library research
- Technology-oriented digital library research & development (R&D)
- Digital library conferences
- Commercial digital library developments
- Criteria for assessing research reports
- Conclusions, challenges

Research orientations

- general division

Basic research

the systematic study aimed at gaining more comprehensive knowledge or understanding of the subject under study, *without a specific or practical application in mind.*
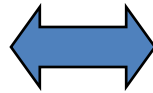
Applied research

the systematic study directed toward the practical application of knowledge and problem solving

refers to scientific study that seeks to *solve practical problems*

It is not that binary & exclusive, but on a continuum from basic to applied

Really, really basic digital library research

- Research oriented toward digital library theories not really done yet
- We do not have theories related to digital libraries
  - or theoretical constructs about any of its components or aspects
- In other words, a theoretical base (or bases) for digital libraries does not exist as yet
  - one of the reasons why we do not have a more systematic research overall
- Some sociological & cultural research started asking basic question

Research orientation in DL

More basic

- Deals mostly with humans & digital libraries
  - users & use of DLs
- But also with contexts - institutions, domains, - roles, evaluation, effects
- *Seeks understanding that may be (hopefully) applied to better designs, applications*

More applied

- Deals mostly with technology & DLs
- But also with many other aspects – e.g. economic, legal, managerial
- *Seeks solution to specific problems*

Setting agendas for digital library research

Most more **basic agendas** were set from bottom up

- from institutions, incl. many libraries, librarians
- imprint of institutional missions, & library & information science community's interests & vision

Most more **applied agendas** were set from top down

- from funding agencies to projects
- imprint of the computer science community's interest & vision

The two are weakly connected, if at all

Emphasis here on

1. Research that dealt with humans & digital libraries

2. Research projects & initiatives that dealt with technology & digital libraries

- Other research areas e.g. context or economics, not treated here

Distribution of digital library research

- Most done on technical aspects,
  - content handling in various media, organization, operations, access tools …

- Relatively well funded in the past

- Much less done on humans and digital libraries
  - still there are a number of studies in a variety of countries

- Relatively meagerly funded in the past

Oh well… Money talks

Human-oriented DL research

- Lot of this research is related to:
  - **human information behavior**
  - **information seeking**
  - **information searching**

in digital environments in general and digital libraries in particular

- Digital libraries provide the base, orientation & context

Where do humans fit?

- Digital libraries support users in informational tasks and needs
  - what are people doing?
  - how are people engaging with a digital library?
  - what were the outcomes?
- But, as yet, we are not completely sure as to
  - what processes & variables to observe
  - what context & effects to include
 - what models, criteria, measures, methodologies to use

Context of human-oriented studies

- Some studies indeed concentrated on human information behavior in digital libraries

- But, most were a part of digital library evaluation
  - users used various features/processes
  - AND evaluation was goal and findings on human information behavior were a byproduct

- Digital library evaluation on the one hand &
  studies of human information behavior in digital libraries on the other hand,
  are mixed together

Classes of variables studied

- in human-oriented digital library research
- as reported in the literature

- **Users** – who uses a digital library? Why?

- **Features** – what features are used?

- **Access** – how is it used?

- **Use** – what content is used? for what?

- **Usability** – ease? effectiveness?

- **Outcomes** – with what results?
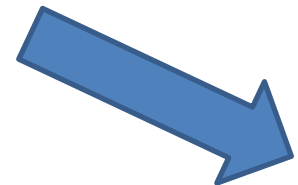
Users – main research questions in various studies

- **population, reasons**
  - who uses a given digital library and why?
- **tasks, queries**
  - what is their nature?
- **knowledge, understanding**
  - what knowledge & learning needed?
- **preferences, expectations**
  - what is preferred? expected?

Features – research questions

- **features**
  - what available features in a digital library are being used?
- **frequency & amount**
  - how much are they used? distribution?
- **patterns**
  - what may be patterns in use of any feature?
- **individual differences**
  - how do individuals differ in use of features?

Access – research questions

- access
  - how? from where? how often? repeated?
- discovery
  - how searched, navigated, browsed, retrieved?
  - what barriers encountered?
- time, effort
  - what does it take to use & find?

Use – research questions

- ## distribution
  - how many items in collection used?

- ## application
  - for what were the items actually used?

- ## social
  - what are social practices associated with digital library use?

Usability – research questions

- **effectiveness**
  - how well are given user tasks accomplished?
- **efficiency**
  - at what cost, effort, time?
- **ease**
  - how easy are interfaces & features to use?
- **barriers**
  - what hindrances? frustrations? lost?
- **learning effort**
  - how much needs to be learned? how? errors?

Outcomes – research questions

- **downloads**
  - what & how much is downloaded?
- **assessments**
  - what value? relevance? usefulness? satisfaction?
- **effects**
  - what impact on activities?
    - e.g. education, scholarly communication
- **quality**
  - what essential attributes support excellence assesments?

Digital library metrics – getting numbers

- *Metrics:* a system of measurement with defined units for measuring – numerical, statistical

- For digital libraries many metrics were used for processes & collections (numbers for: access, use, downloads, size …)

- But no standardized set of metrics yet
  - some are used more universally, e.g. Assoc. Res. Libr. statistics, LibQUAL+ (quantitative data on service quality perceptions incl. instruments)

- Some metrics are now supplied by vendors – e.g. usage

- At issue: what do you measure? how do you measure? – hard questions!

Some interesting results

- Users
  - half are repeated users
- Features
  - many not used
- Access
  - varies by nature of desired materials
  - users vary in skills & understanding

results …

- Use
  - 20–80 rule seems valid – a small proportion of collection (20% or less) is used (80% of times or more)
  - downloads are significant
- Usability
  - many barriers to use
  - effectiveness varies greatly
  - frustration is easily triggered

results …

- Outcomes
  - many queries do not lead to viewing
  - text materials not appealing
 - valuing varies with age

Deriving generalizations

- Suggest here a model for use in studying humans & digital libraries
  - shows: what is to be studied

BTW:

- Models are not theories
  - theories explain & predict; models enumerate
  - theories are about why and how; models are about what is involved or going on
  - theories guide; models provide structure

A model for studying human inf behavior in digital libraries – what to study?

users, tasks, contexts

features

access

seeking inf

use

usability

d-library

outcomes
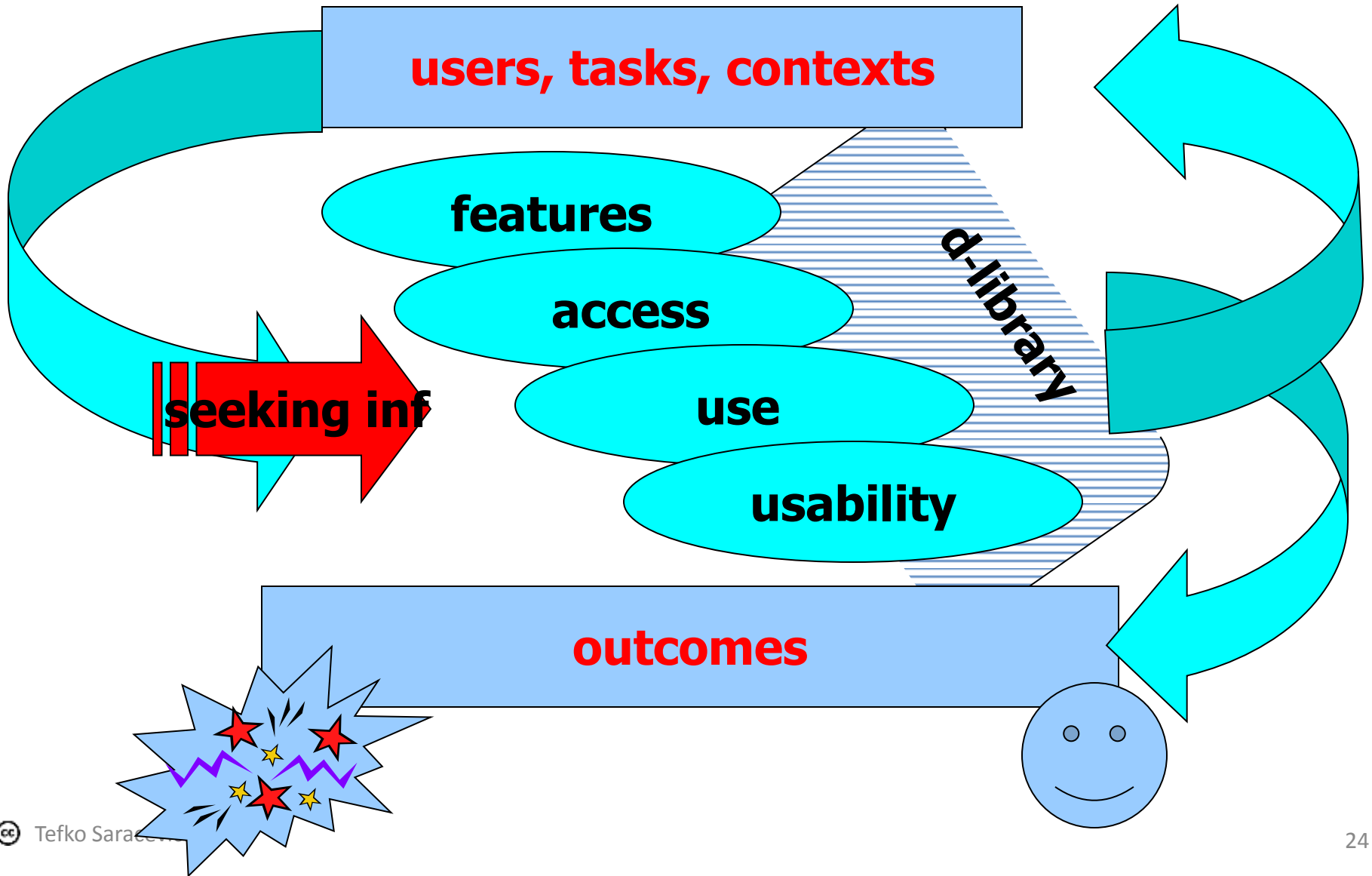
Technology-oriented digital library research & development (R&D)

- Technical problems are much larger and complex than anticipated
  - digital libraries are complex systems
  - deal with a variety of forms of artifacts of human knowledge – from texts to music
  - involve language and concepts with all idiosyncrasies
  - and with problems inherited in digital media
  - and with information explosion – large scale

Recognition

- Of technical challenges & opportunities
  - & of importance of digital libraries to society – part of information society
- Brought about a number of R&D initiatives & projects with technical orientation
  - where "technical" also involved developing digital libraries as support for defined communities, domains

Initiatives in the US

– historical beginning of digital library research

- National Science Foundation lead Digital Library Initiatives phase 1 & 2
  - DLI 1:1994-1998; 6 projects; $30 mill.
  - DLI 2:1999-2006; 36 projects; ~$55 mill.
- Technological or developmental in nature
  - addressing difficult problems of applying technology
- Considerable impact in creating interest & communities – persists

Dig Libr Initiatives 1 & 2

- At the time there was a huge international commotion about DL initiatives & projects
  - but historical record of these important projects & research phase is disappearing
  - NSF fact sheet describes some projects
  - but interestingly DLI 1 & 2 sites disappeared from NSF site – these are no more there:
  - www.dli**2**.nsf.gov/ ; International projects: www.dli**2**.nsf.gov/intl.html
- DLI projects can be gleamed from
  - project features in D-Lib Magazine
  - or Blue Book

Historical examples of research projects

- [Alexandria Digital Library](#) (UCSB)
  - georeferenced materials
- [Informedia Digital Video Library](#) (Carnegie Melon U)
- [Searching](#) technical documents on the web (UIUC)
- NSF supported a [Stanford](#) DLI 1project led by Hector Garcia-Molina & Terry Winograd:
  - infrastructure and services needed for collaborative DL functions
  - and (as already mentioned) Google came out of that project (at least in part, as already mentioned) – done by Larry Page and Sergey Brin (Google founders) while students on that project
  - NSF: [On the origins of Google](#)

Further initiatives:

- # NSF National Science Digital Library (NSDL)

  – "to provide quality digital resources to the science, technology, engineering, and mathematics (STEM) education community, both formal and informal, institutional and individual."

    - much of it oriented toward K-12 education
    - 2000 up to present; over 300 projects; over $150 mill.
    - basically developmental in nature in various STEM subjects, with some addressing larger issues, such as metadata

National Science Digital Library initiative

- Consist of projects in variety of subjects where many collections are housed & integrated

- Browse by collection shows high diversity e.g.
  - DL for Earth System Education
  - Physics Teaching Web Advisory
  - Instructional Architect (for teachers, students)

- **BIG** issue: sustainability – NSF funded development, but what now after grants are over?

  - Who will keep them up?

  - A number are going on, but some already disappeared

European Community

- <u>Project DELOS</u> on Excellence in Digital Libraries
  - "integrating & coordinating the ongoing research efforts of the major European teams working in Digital Library-related areas." 2004-2009
- Continued with projects under <u>Digital Library Interoperability, Best Practices and Modelling Foundations</u> – six working groups. 2009-2011
- Digital Agenda for Europe: Digital Libraries Initiative – among others, development of <u>Europeana</u>
  - part of <u>Digital Agenda for Europe</u> – ongoing till 2020

Digital library conferences

Major vehicles for reporting DL research:

- Joint Conference on Digital Libraries (JCDL) (joint between ACM & IEEE) - US

- International Conference on Theory and Practice of Digital Libraries (TPDL) - Europe
  - before 2011 it was named European Conference on Research and Advanced Technology for Digital Libraries (ECDL)

- International Conference on Asian Digital Libraries (ICADL)

Topics of papers at JCDLs

- Mostly technical – such as: architecture; data
- But also covered topics as: preservation; bibliographic networks; metadata; books & reading; user behavior …
- Papers generally limited to max 10 pages
- All have sessions with posters – often groundbreaking developments & issues

And a small conference

(Tooting my own horn – I was a co-director till 2014)

Libraries in the Digital Age (LIDA)

- Held in Zadar, Croatia
  - biennially
- 2012: ~160 participants from 24 countries; many students
- 2014: ~140 participants from 22 countries
- Next: in 2016

Themes:

- LIDA 2012:
  - CHANGES in the world of library services: Evolution and innovation
  - CHANGES in the world of electronic resources: Information and digitization
- LIDA 2014
  - Qualitative methods in assessing libraries, users, & use: applications, results
  - Altmetrics - new methods in assessing libraries, users, & use: applications, results

Commercial digital library R&D

- Doesn't really belong under Research, but a lot of Development
  - major sources of innovation
- Mostly library software & systems
  - many areas
- Follow what research community does to some extent, but have own developmental & innovation teams

- Many companies, vendors
- Libraries completely dependent on them
- Librarians prepare elaborate Requests for Proposal (RFP) for companies to bid
- Evaluate software, hardware, $$$$
- Need to keep up with advances

Examples of commercial developments

- Ringgold
- Ex Libris
- LibLime open source
  - Koha
- SirsiDynix

- Many mergers, acquisitions, consolidations
  - volatile commercial area

Some good news & information sources:

- Library Technology Guides
  by Marshall Breeding, Vanderbilt U
- Stephen's Lighthouse
  by Stephen Abram, VP SirsiDynix
- Information Today
  print & online newspaper

Criteria for assessing research articles in library literature - example

Hildreth & Aytac (2007) survey - 206 articles, 23 journals

- Well organized?
- Research question/problem/aim stated at the beginning?
- Literature review?
- Data collection explained?
- Data-supported presentation of results?
- Visual presentation of data findings?
- Validity/reliability issues addressed?
- Limitations of study mentioned?
- Future research recommendations discussed?

Can be used as criteria for assessing any research

Criteria for assessing quality of research submissions to journals - example

Information Processing & Management (IP&M)

- Reviewers guidelines (Elsevier)

1. Appropriateness for IP&M (to Aims & scope)
2. Originality and value for future work
3. Quality of methodology and treatment
4. Adequacy of citations and recognition of previous work
5. Validity of claims and interpretations
6. Organization, style and clarity of presentation

Conclusions

- R&D in digital libraries is global
- Two major tracks
  - technological – **BIG** & growing
  - human – small, not funded much
- Of major importance for DL practice
- Sometimes hard to follow, but must be followed
- Many librarians active participants
- Systems librarians involved most
- How about getting involved???

Challenges: connections

- Enhancing the research – practice dialog
  - establishing connections

- Using, translating results
  - between research & practice

Research

Digital libraries

Practice

Lecture on Wordle