Evaluating Search in Digital Libraries

Presentation at the National Science Digital Library
Annual Meeting

November 8

Anne Diekema

SYRACUSE UNIVERSITY
School of
Information
Studies

Evaluating digital libraries is a little like judging how successful is a marriage.

(Marchionini, 2000)

Difficulties with digital libraries

- In a study looking at 80 digital library evaluation reports Saracevic (2004) found the following common problems

  - Users do not fully understand digital libraries
  - Users hold a different conception of a digital library from the library builders
  - Users lack familiarity with the capabilities of the library
  - Users do not know what is contained in a digital library
  - User have limited capabilities interacting with the library

I am your Venus

Users are from Venus and digital libraries are from Mars

(Saracevic, 2004)

www.solarviews.com/cap/venus/venus1.htm

Search

- The main interaction of users with the digital library is through search
- The Google effect
- Systemic approach (Muresan, 2005)
  - Return as many relevant documents as possible
  - Filter out non-relevant documents
- Cognitive approach
  - Support the user's exploration of the problem domain and aid in the completion of their tasks
- Each approach requires different type of evaluation
  - Systemic approach => system evaluation
  - Cognitive approach => user evaluation

Two research paradigms

- The physical paradigm
  - Analogy of retrieval system being a physical system
  - Primary focus: physical or artefactual identity of the system
  - 1953 – Cranfield tests; Empirically grounding system quality
  - Main research focus: development of retrieval models and techniques through controlled experimentation
- The cognitive paradigm
  - Primary focus: people – cognitive or human aspects of the system
  - Main research focus: development of techniques for modeling the cognitive world of the user as part of the retrieval interaction
- (Ellis, 1992)

System Evaluation

- Requires test collection
- Combination of documents, test queries, and their relevance judgments
- Document relevance determined in advance
- Experiments
  - use queries to retrieve documents from test collection
  - calculate performance based on the retrieved documents and the relevance judgments
  - evaluation measures reflect how well a system does at finding relevant documents and ignoring irrelevant documents (Van Rijsbergen, 1981)

Relevance, Recall and Precision

- Derived measures: recall, precision, fallout, omission factor, noise, specificity
  - Recall — ratio of the number of relevant retrieved documents to total number of relevant documents in the collection
  - Precision — ratio of the number of relevant retrieved documents to total number of retrieved documents
- Employment of people (human judgments) as measuring instruments – people make for fickle instruments
- Relevance assumptions unrealistic

User evaluation

- Qualitative approaches:
  - Cognitive approaches
  - Behavioral approaches
  - Affective approaches
- User satisfaction
- System walk-thrus, focus groups, interviews, think aloud

Conclusions

- To evaluate Search in NSDL we need to evaluate the system itself as well as the user in context using the library
- System evaluations can precede user evaluation
- User evaluation more time consuming but the ultimate barometer of what is going well and what needs to be changed