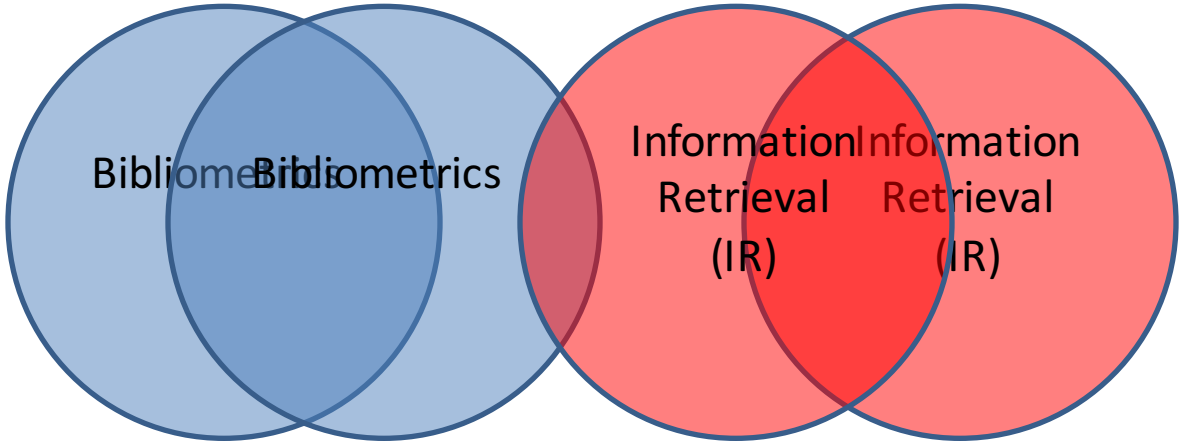
Bibliometrics & Information Retrieval Overview

Dietmar Wolfram
University of Wisconsin-Milwaukee

ASIST 2016

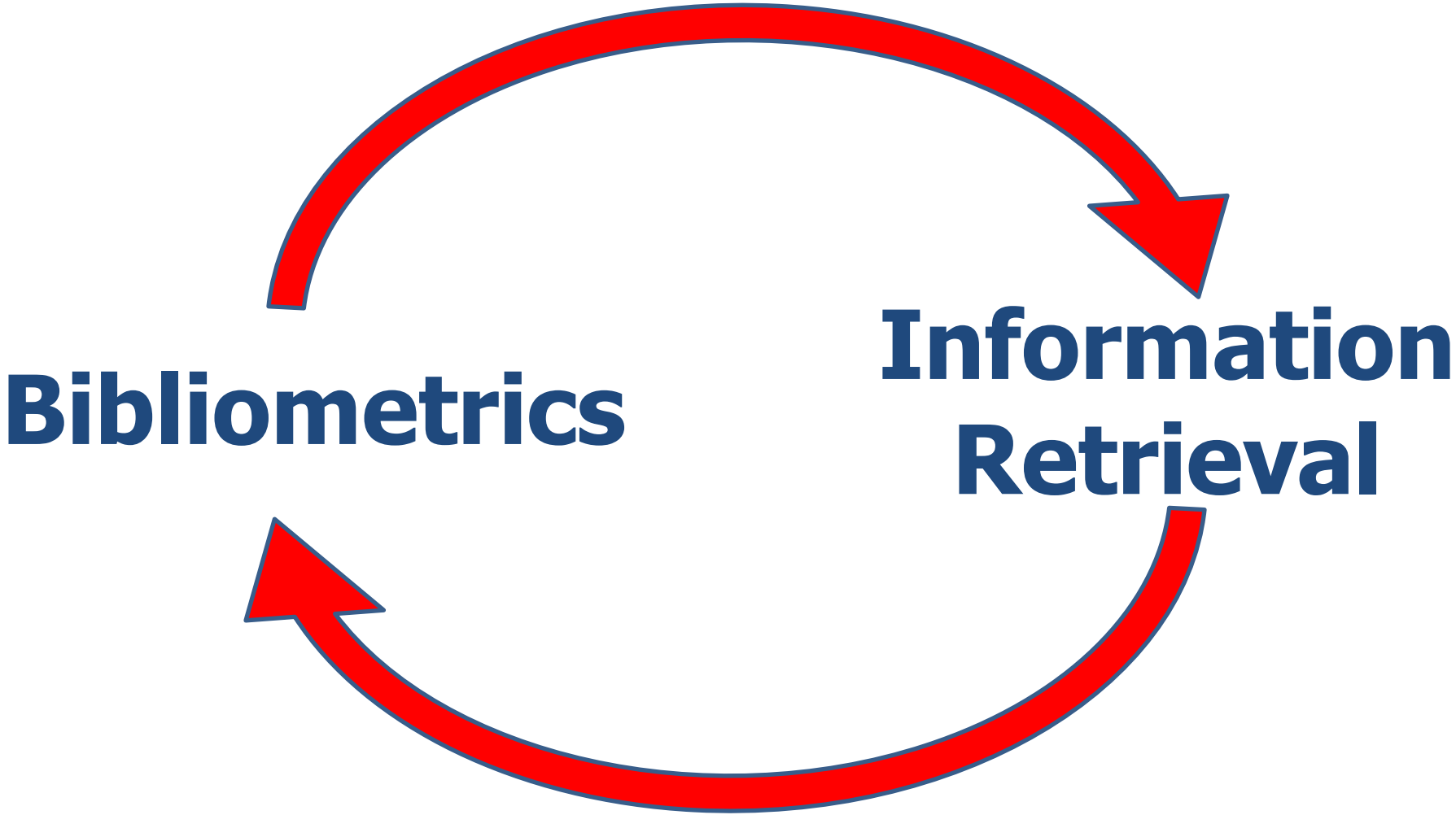
Overview

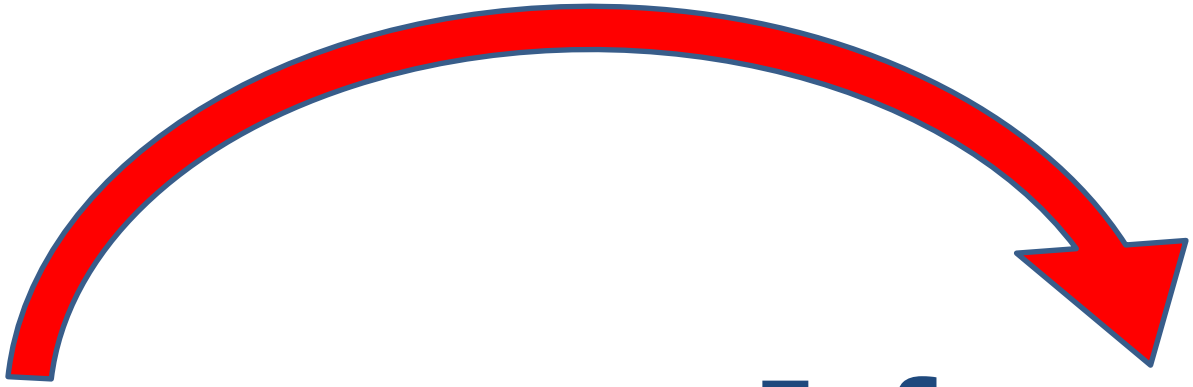


Introduction

- The intersection of two key areas of information science offers many areas for research

- Recent BIR workshops demonstrate growing interest in the synergies between the two

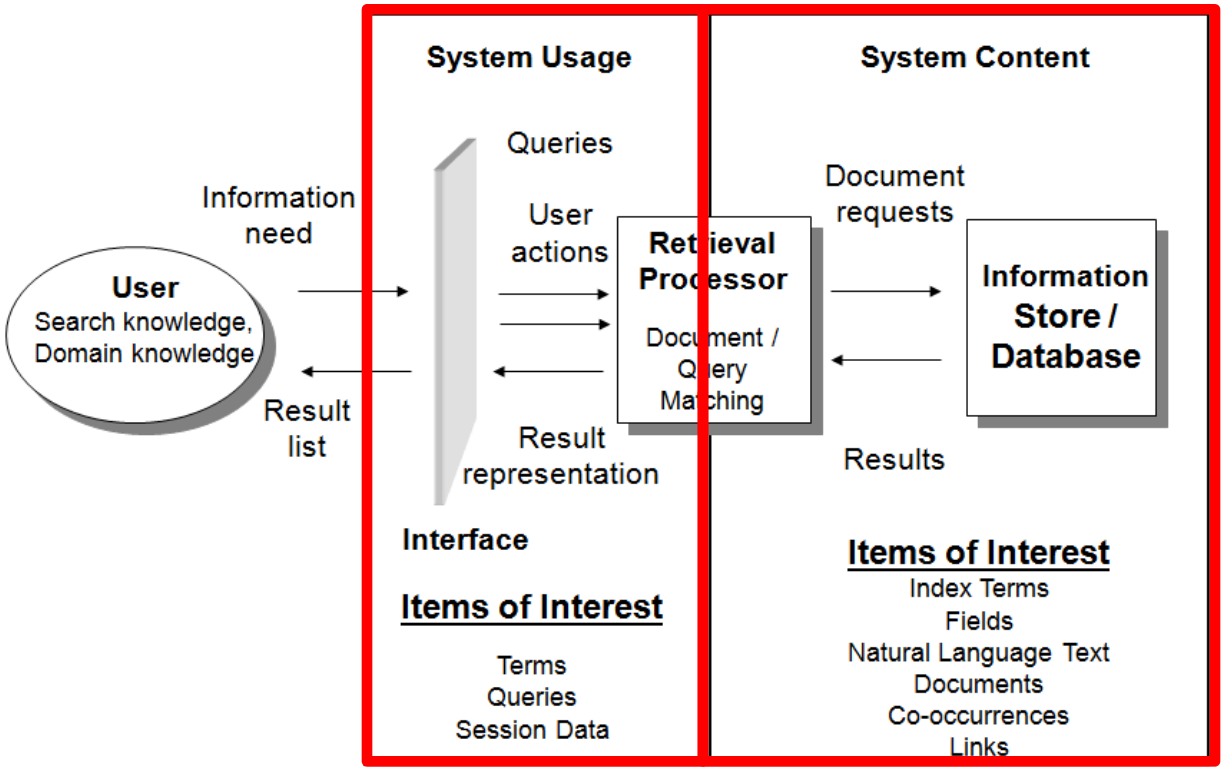

Bibliometrics
Information Retrieval



Bibliometrics

Information Retrieval

IR Processes & Associated Data



Adapted from Wolfram, D. (2003). *Applied informetrics for information retrieval research*. Westport, CT: Libraries Unlimited.

Areas of Application

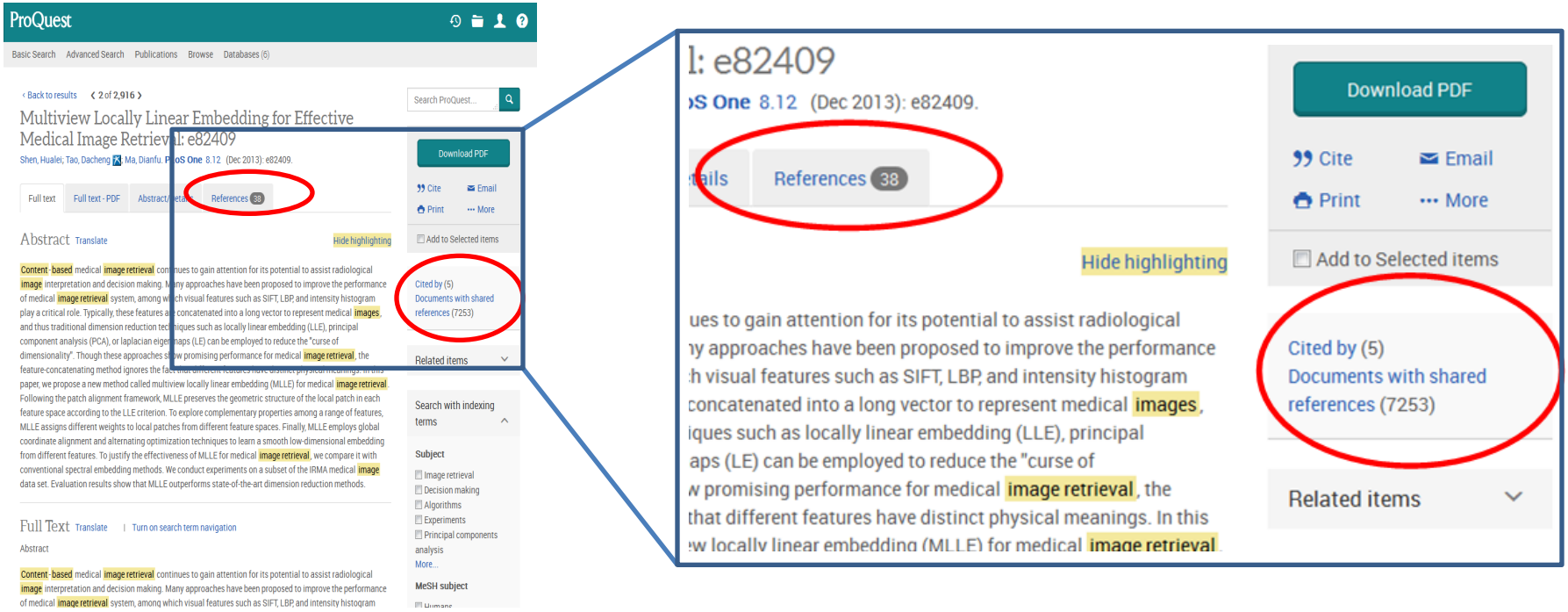
- **Modeling IR processes**
  - System indexing & retrieval
  - IR system simulation

- **IR & allied system design & evaluation**
  - Using graph-based approaches / link analysis (co-authorship, citations, hyperlinks)
    - Ranking results
    - Supporting browsing & expanding results

IR System Usage

- Content Use
  - Website visitation
  - Document requests
- User search characteristics
  - Terms
  - Queries
  - Sessions (search and browsing actions)

Linking Citing & Cited Documents

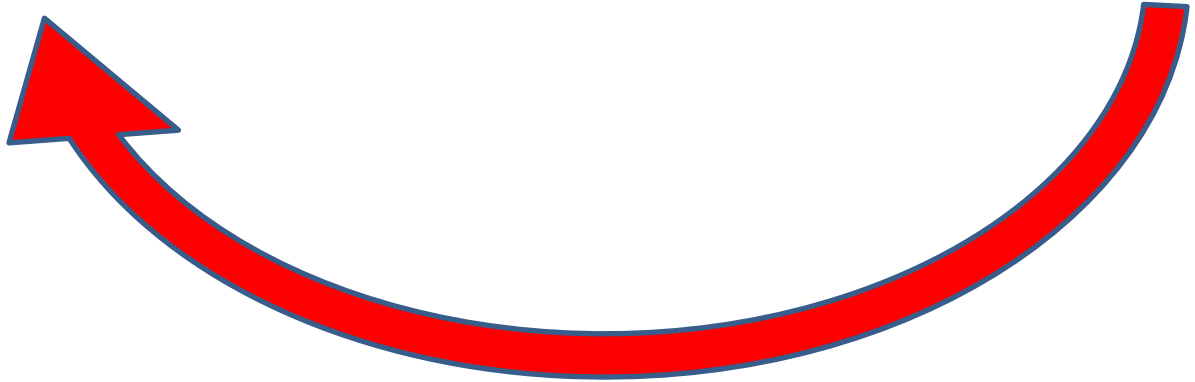


Ranking Documents

- HITS (Kleinberg, 1997)
- PageRank (Page et al., 1999)
- Hw-rank (Bar-Ilan & Levene, 2015)
- Bradfordizing & author centrality (Mutschke & Mayr, 2015)
- Article-level Eigenfactor (Wesley-Smith, Bergstrom, & West, 2016)

Bibliometrics

Information Retrieval



Reciprocal Contributions

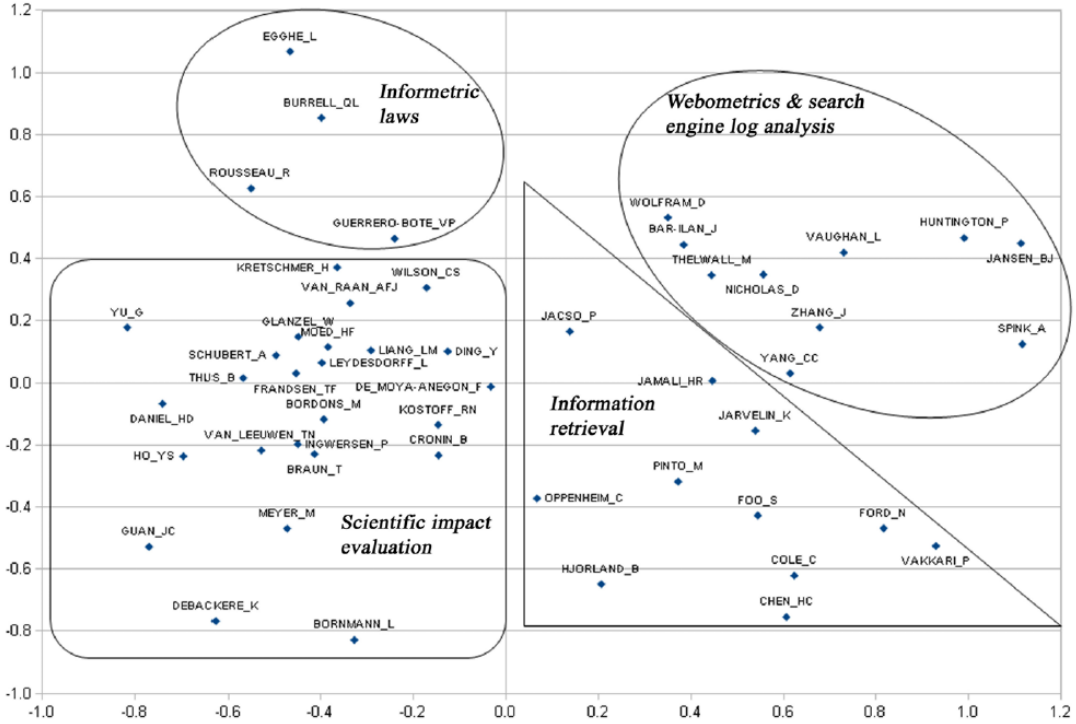
- With growing datasets, new ways to store, process and display data are needed
- IR frameworks provide tools & approaches for -metrics researchers
  - Database design for bibliographic datasets
    - Relational & graph-based DBMSs, IR software & toolkits
  - Application of vector space & probabilistic IR models to compare data

PageRank Comes Full Circle

Language-based Methods & -metrics Research

- Citations & collaborations form the foundation of traditional comparative analysis
- Downside: No link ⇒ No relationship
- Language can expand relationship possibilities
  - Term co-occurrence
  - Topic modeling
  - Identifying hidden patterns with text mining

Author-Topic Modeling for Author Research Relatedness



An A-T model produced more coherent groupings of prolific authors in information science than co-citation analysis

Lu, K., & Wolfram, D. (2012). Measuring author research relatedness: A comparison of word-based, topic-based and author co-citation approaches. *Journal of the American Society for Information Science and Technology,* 63(10), 1973-1986.

Text Mining

- Can be combined with bibliometric methods
  - Citation mining for user research profiling (Kostoff et al., 2001)
  - Clustering of scientific fields (Janssens, 2007)
  - Knowledge structure of bioinformatics (Song & Kim, 2013)

- Text mining techniques are integrated into some bibliometric mapping software, including
  - VOSviewer - http://www.vosviewer.com/
  - CiteSpace - http://cluster.cis.drexel.edu/~cchen/citespace/

Future Directions

- Complexities of bibliometric datasets lend themselves to IR techniques
  - Resulting “big data” require data and text processing or mining techniques to identify overt & hidden patterns
- Topic modeling and other text-based methods show great promise in providing complementary approaches to citation & co-authorship data
  - Computational overhead to train models is still high
- Need for better evaluation methods for visualization outcomes

Thank you