Text Retrieval and Search Engines
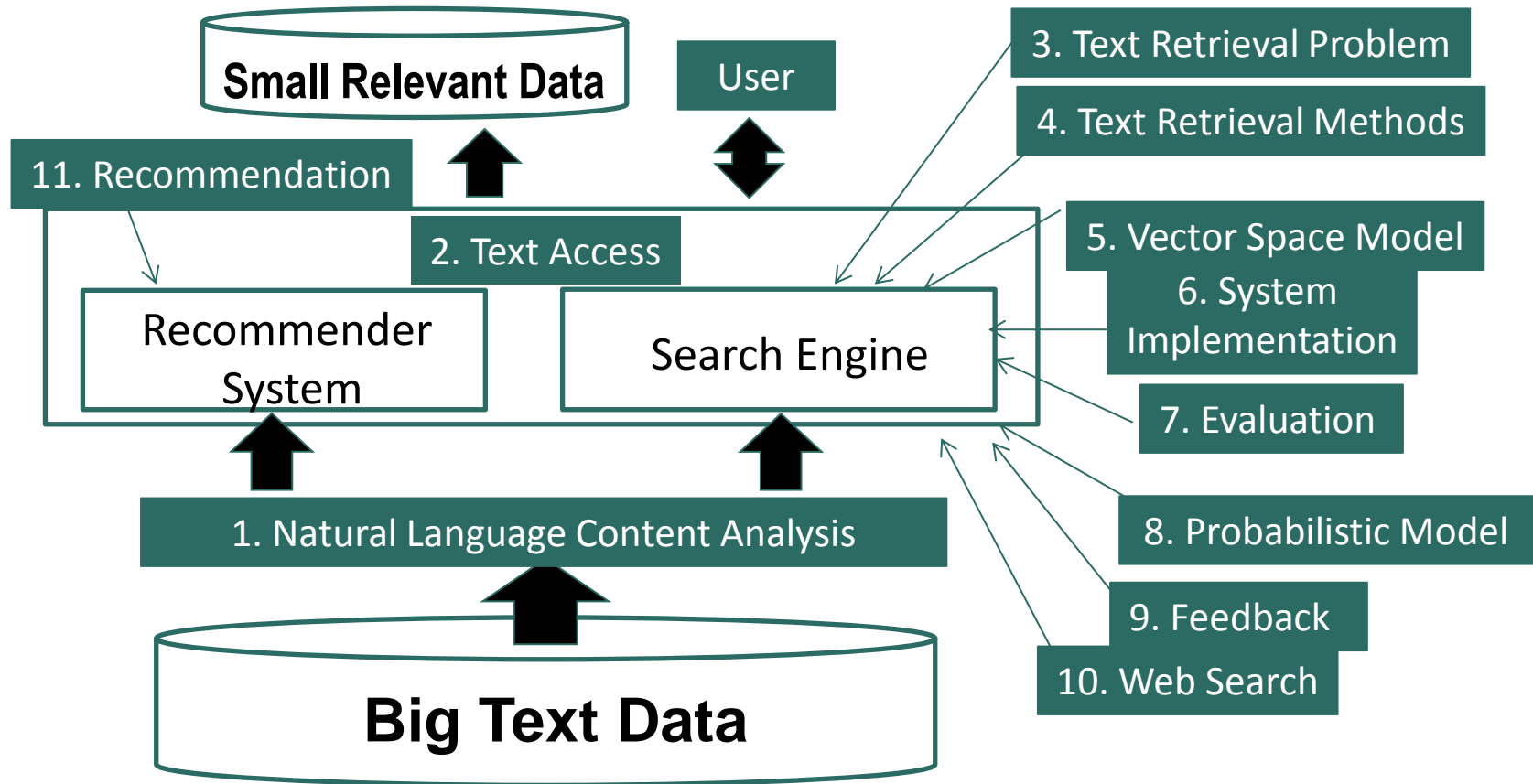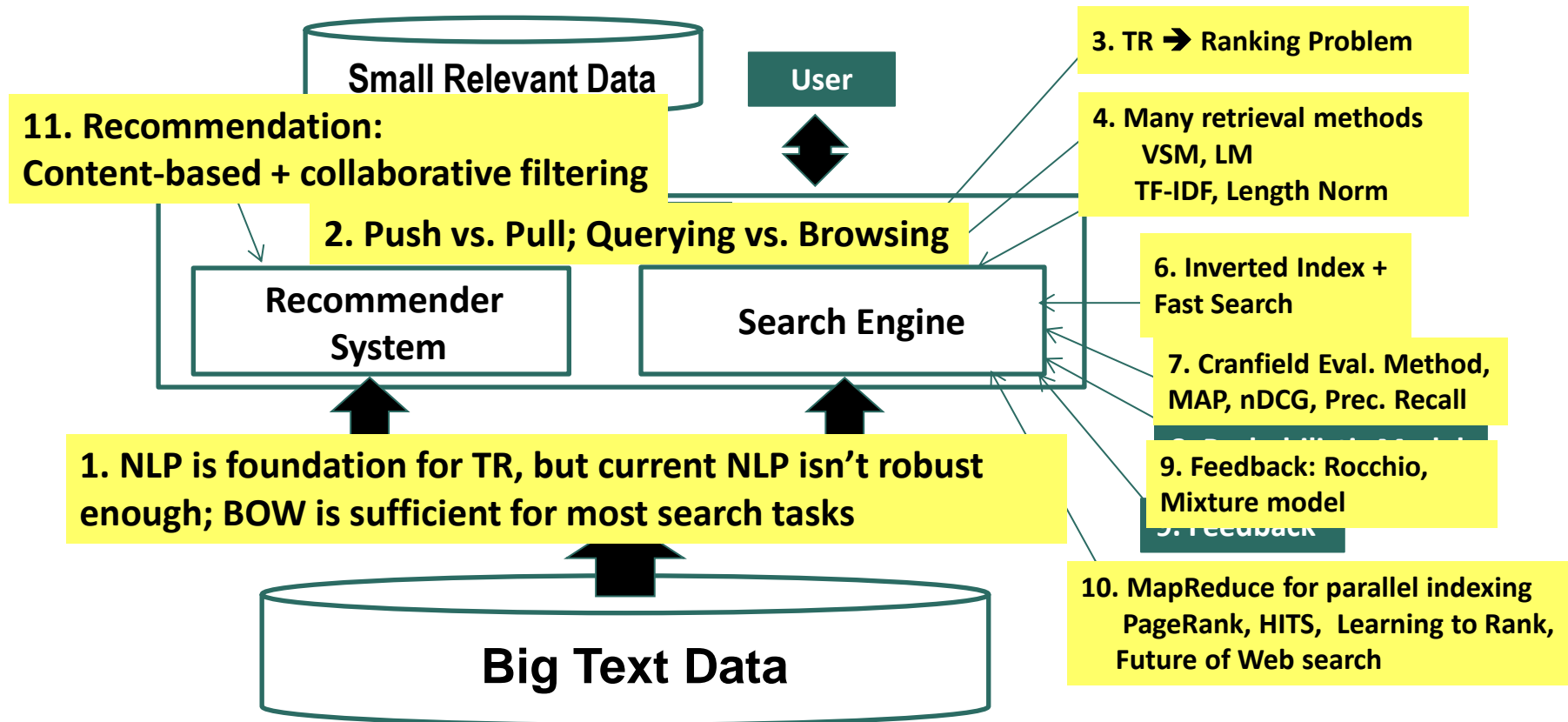
Course Summary

ChengXiang "Cheng" Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign

Course Summary: Major Topics Covered

Key High-Level Take-Away Messages

Small Relevant Data

User

3. TR ➔ Ranking Problem

11. Recommendation:
Content-based + collaborative filtering

4. Many retrieval methods
VSM, LM
TF-IDF, Length Norm

2. Push vs. Pull; Querying vs. Browsing

Recommender System

Search Engine

6. Inverted Index + Fast Search

7. Cranfield Eval. Method, MAP, nDCG, Prec. Recall

1. NLP is foundation for TR, but current NLP isn't robust enough; BOW is sufficient for most search tasks

9. Feedback: Rocchio, Mixture model

Big Text Data

10. MapReduce for parallel indexing PageRank, HITS, Learning to Rank, Future of Web search

Search User Interface,
by Marti Hearst,
Cambridge University Press, 2009
http://searchuserinterfaces.com/

Additional Readings

- Synthesis Digital Library has many excellent short books/long tutorials on relevant topics: http://www.morganclaypool.com/page/ForthcomingSynthesisLectures
  - Information Concepts, Retrieval, and Services: http://www.morganclaypool.com/toc/icr/1/1
  - Human Language Technology:
    http://www.morganclaypool.com/toc/hlt/1/1
  - Artificial Intelligence & Machine Learning: http://www.morganclaypool.com/toc/aim/1/1
- Journals: ACM TOIS, IRJ, IPM, …
- Conferences: SIGIR, CIKM, ECIR, WSDM, WWW, KDD, ACL, …

For more info, check out: http://timan.cs.uiuc.edu/resources

Main Techniques for Harnessing Big Text Data:
Text Retrieval + Text Mining