

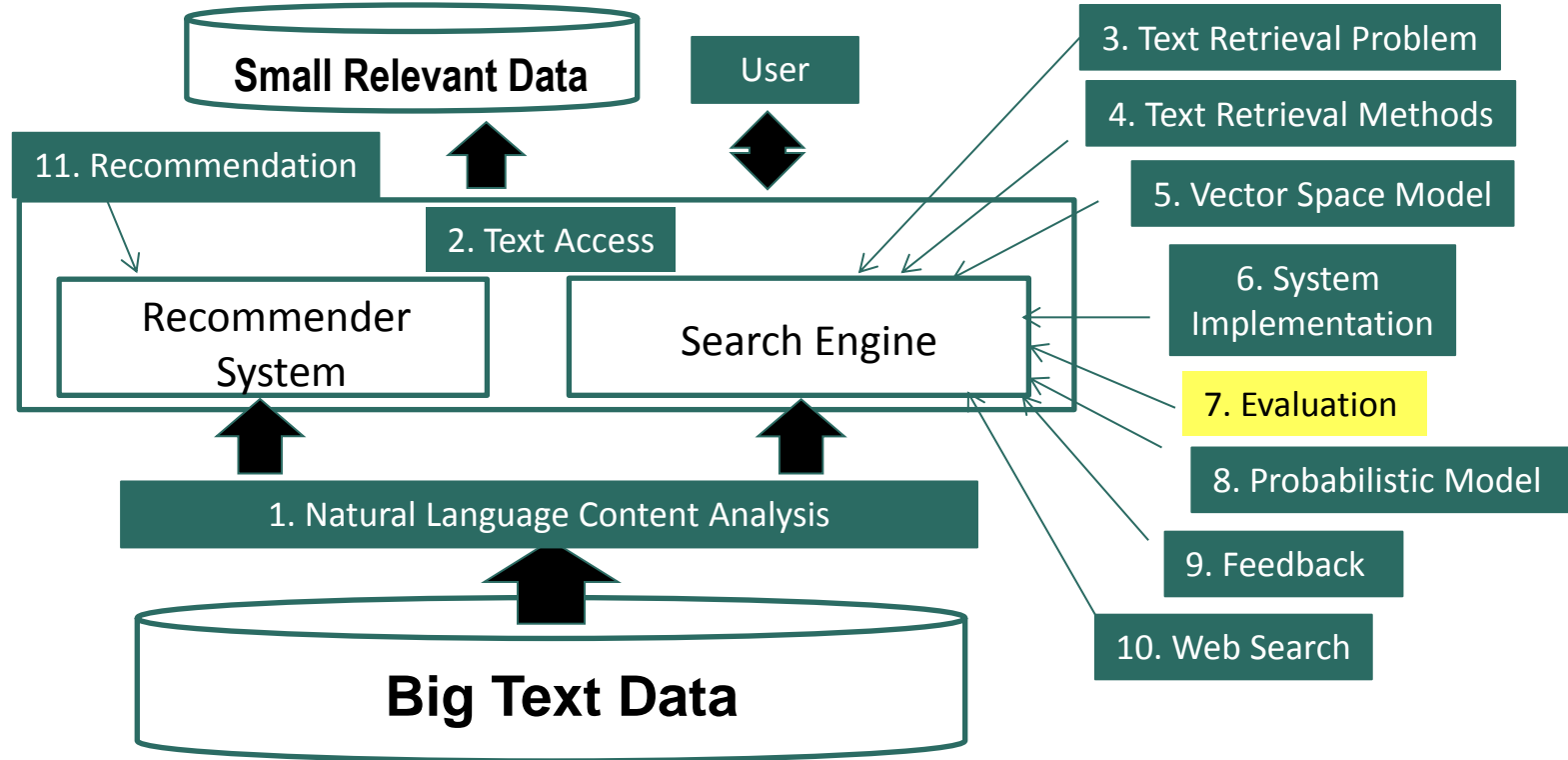


# Text Retrieval and Search Engines

Evaluation of TR Systems: Evaluating a Ranked List Part 2

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

# Evaluation of TR Systems: Evaluating a Ranked List



# Mean Average Precision (MAP)

- Average Precision:
  - The average of precision at every cutoff where a new relevant document is retrieved
  - Normalizer = the total # of relevant docs in collection
  - Sensitive to the rank of each relevant document
- Mean Average Precisions (MAP)
  - **MAP** = arithmetic mean of average precision over a set of queries
  - **gMAP** = geometric mean of average precision over a set of queries
  - Which is better: MAP or gMAP?

# Special Case: Mean Reciprocal Rank

- When there's only one relevant document in the collection (e.g., known item search)
  - Average Precision = Reciprocal Rank =  $1/r$ , where  $r$  is the rank position of the single relevant doc
  - Mean Average Precision → Mean Reciprocal Rank
  - Why not simply use  $r$ ?

# Summary

- Precision-Recall curve characterizes the overall accuracy of a ranked list
- The **actual** utility of a ranked list depends on how many top-ranked results a user would examine
- Average Precision is the standard measure for comparing two ranking methods
  - Combines precision and recall
  - Sensitive to the rank of **every** relevant document

What if we have multiple levels of relevance judgments?