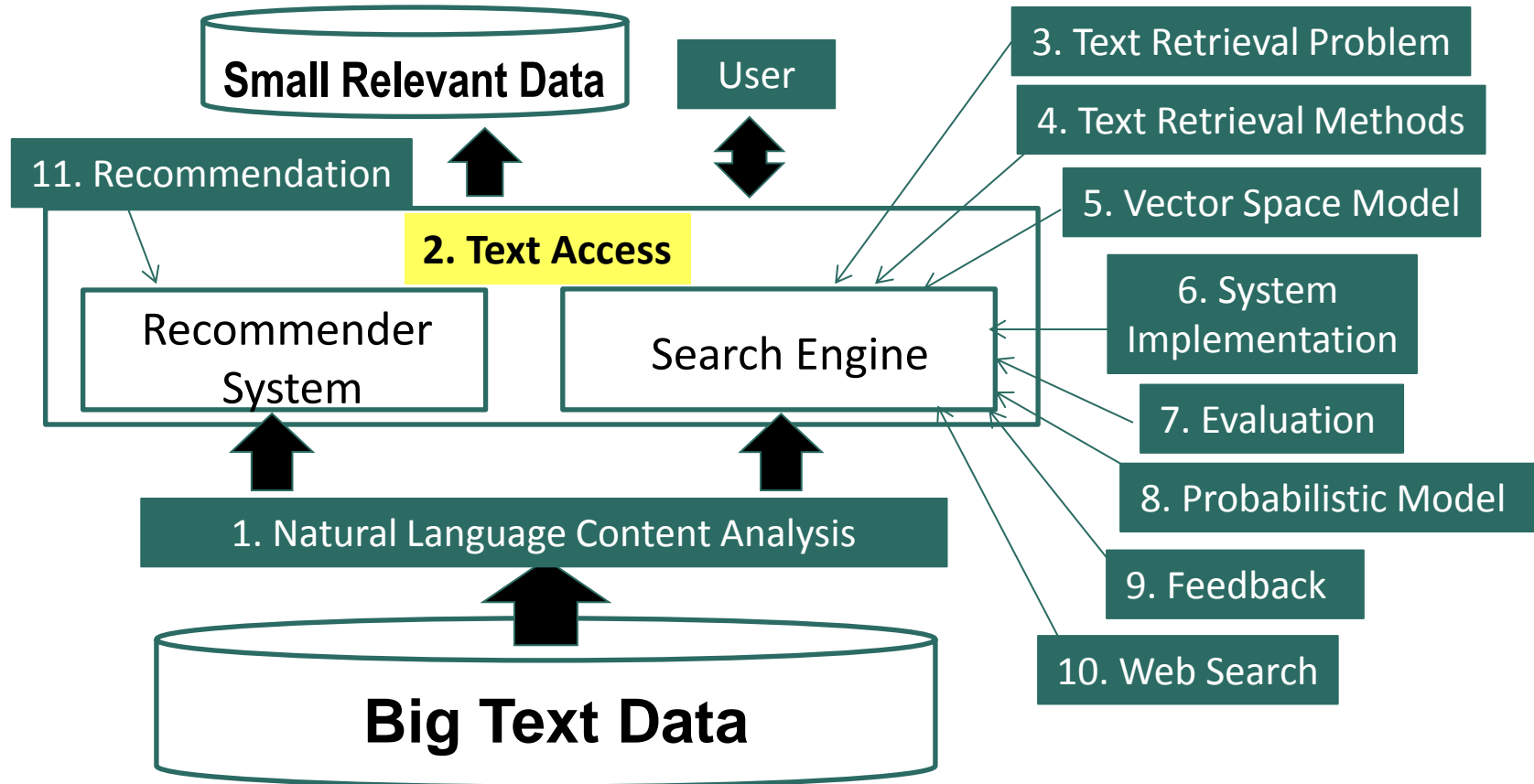# Text Retrieval and Search Engines

Text Access

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

Course Schedule

Access to Relevant Text Data

How can a text information system help users get access to the relevant text data?

- Push vs. Pull

- Querying vs. Browsing

Two Modes of Text Access: Pull vs. Push

- **Pull** Mode (**search engines**)
  - Users take initiative
  - Ad hoc information need
- **Push** Mode (**recommender systems**)
  - Systems take initiative
  - Stable information need or system has good knowledge about a user's need
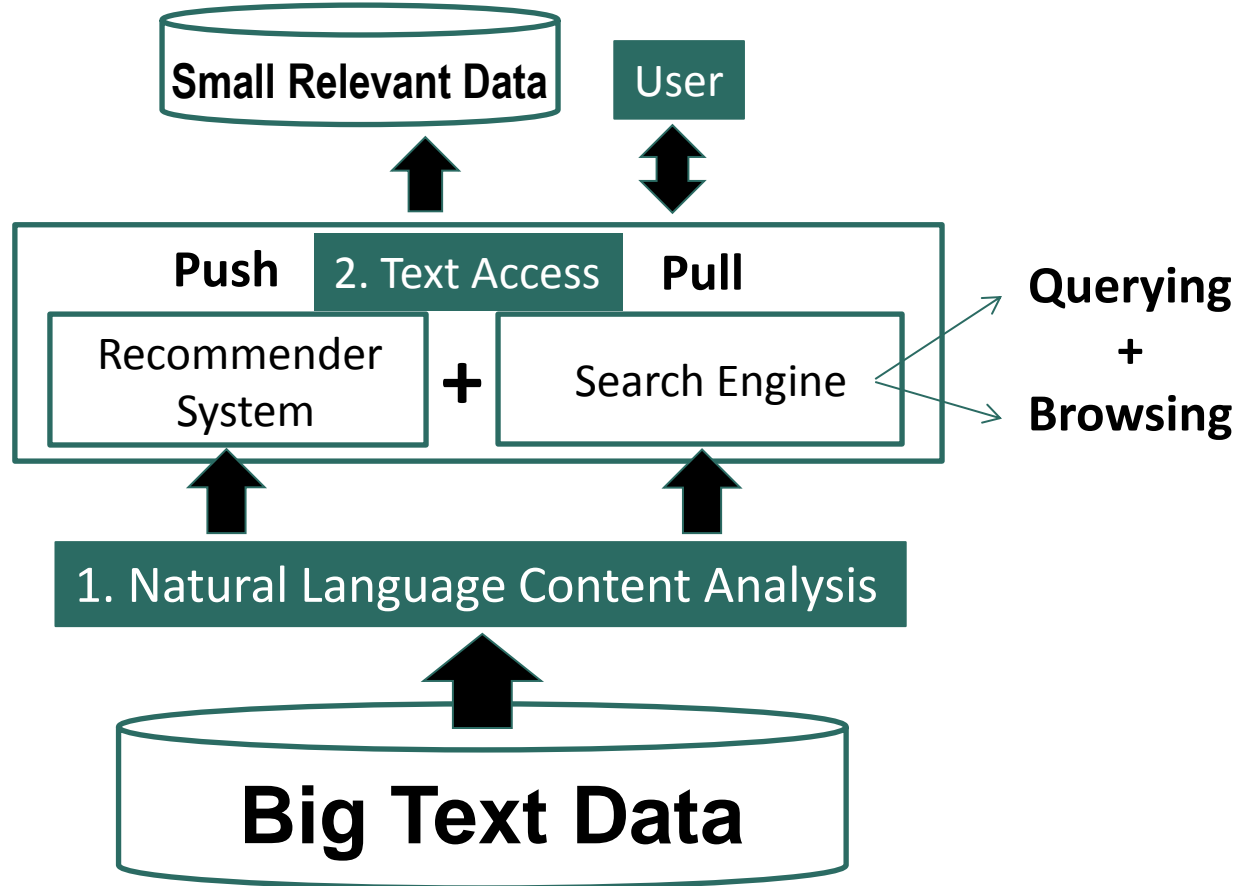
Pull Mode: Querying vs. Browsing

- Querying
  - User enters a (keyword) query
  - System returns relevant documents
  - Works well when the user knows what keywords to use
- Browsing
  - User navigates into relevant information by following a path enabled by the structures on the documents
  - Works well when the user wants to explore information, doesn't know what keywords to use, or can't conveniently enter a query

Information Seeking as Sightseeing

- Sightseeing: Know address of an attraction?
  - Yes: take a taxi and go directly to the site
  - No: walk around or take a taxi to a nearby place then walk
- Information seeking: Know exactly what you want to find?
  - Yes: use the right keywords as a query and find the information directly
  - No: browse the information space or start with a rough query and then browse

Summary

Small Relevant Data

User

Push 2. Text Access **Pull**

Recommender System **+** Search Engine

Querying + Browsing

1. Natural Language Content Analysis

Big Text Data

Additional Reading

N. J. Belkin and W. B. Croft. 1992. Information filtering and information retrieval: two sides of the same coin?. *Commun. ACM* 35, 12 (Dec. 1992), 29-38.