

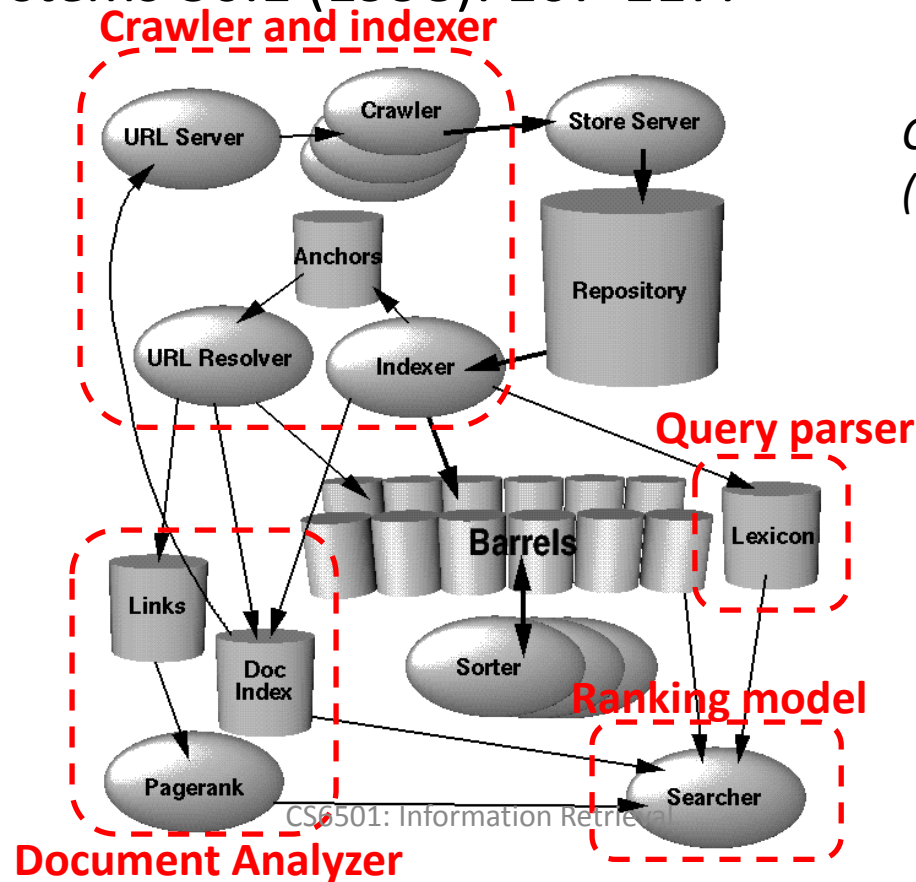
# Search Engine Architecture

Hongning Wang

CS@UVa

# Classical search engine architecture

- ***“The Anatomy of a Large-Scale Hypertextual Web Search Engine”*** - Sergey Brin and Lawrence Page, *Computer networks and ISDN systems* 30.1 (1998): 107-117.



Citation count: 12197  
(as of Aug 27, 2014)

User input

Result display

Query parser

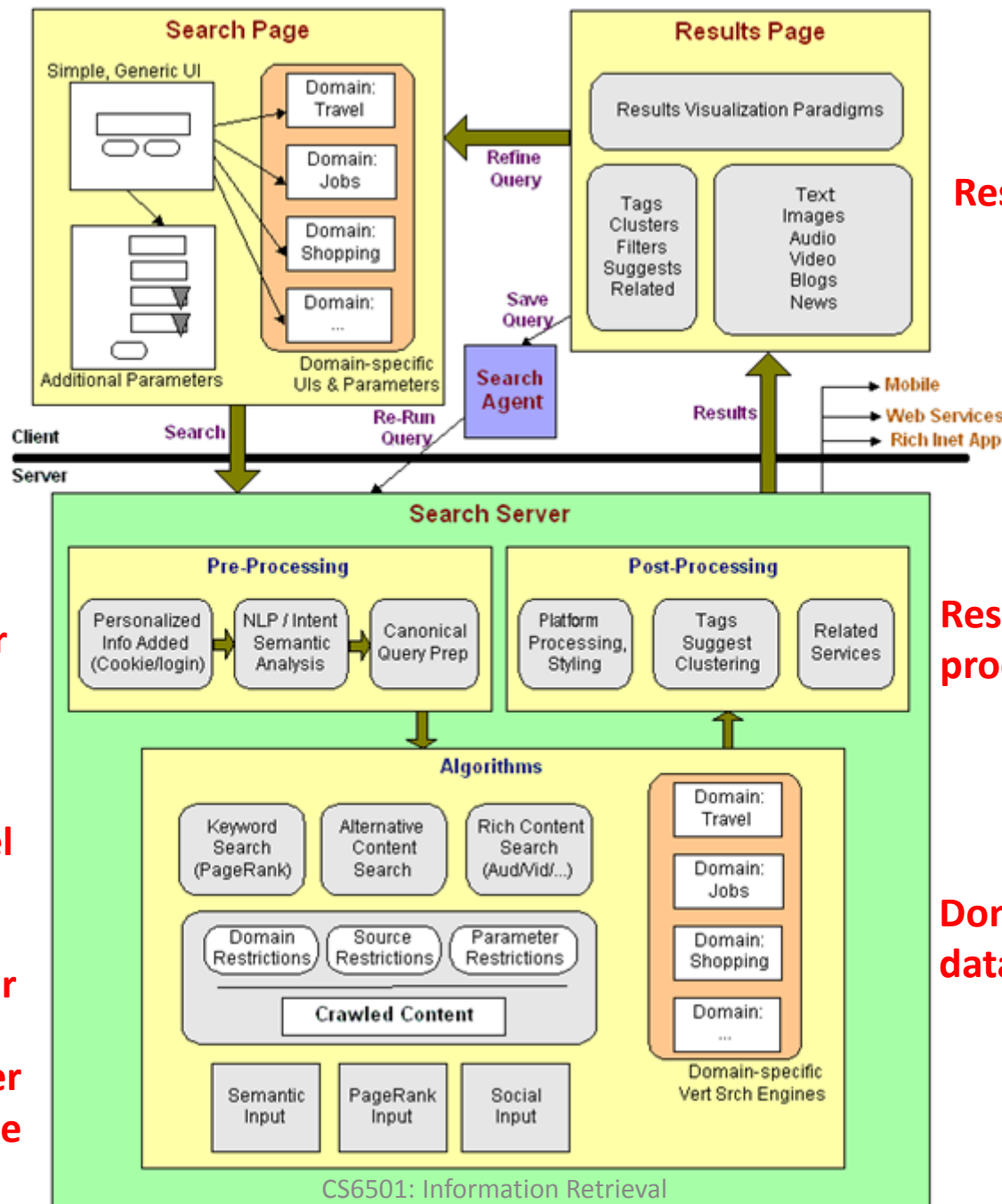
Ranking model

Crawler & Indexer

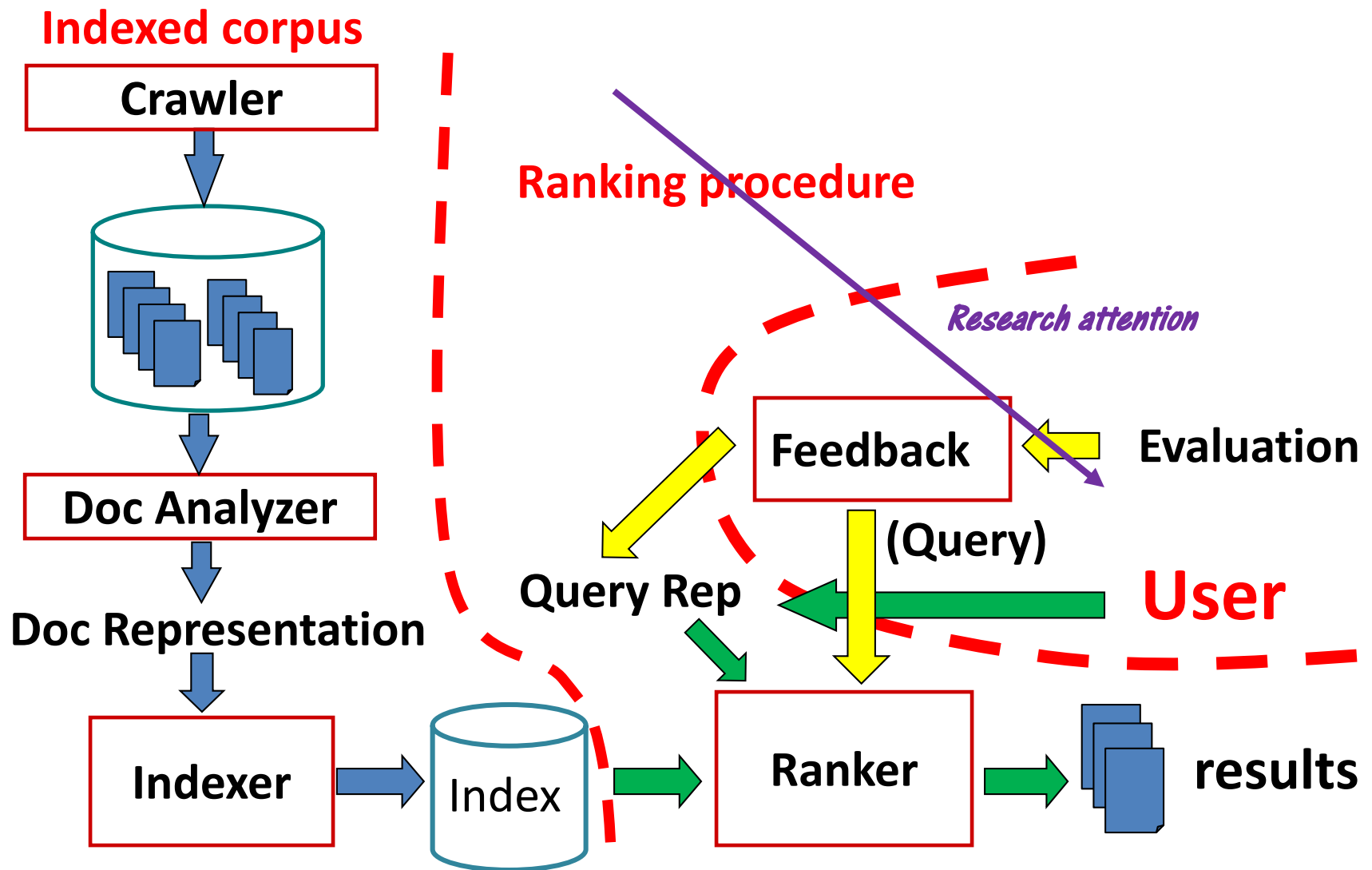
Document analyzer  
& auxiliary database

Result post-  
processing

Domain specific  
database



# Abstraction of search engine architecture



# Core IR concepts

- Information need
  - “*an individual or group's desire to locate and obtain information to satisfy a conscious or unconscious need*” – wiki
  - An IR system is to satisfy users' information need
- Query
  - A designed representation of users' information need
  - In natural language, or some managed form

# Core IR concepts

- Document
  - A representation of information that potentially satisfies users' information need
  - Text, **One sentence about IR - “rank**
- Relevance ***documents by their relevance to***
  - Related ***the information need***  
information need
  - Multiple perspectives: topical, semantic, temporal, spatial, and etc.

# Key components in a search engine

- Web crawler
  - A automatic program that systematically browses the web for the purpose of Web content indexing and updating
- Document analyzer & indexer
  - Manage the crawled web content and provide efficient access of web documents

# Key components in a search engine

- Query parser
  - Compile user-input keyword queries into managed system representation
- Ranking model
  - Sort candidate documents according to its relevance to the given query
- Result display
  - Present the retrieved results to users for satisfying their information need



# Key components in a search engine

- Retrieval evaluation
  - Assess the quality of the return results
- Relevance feedback
  - Propagate the quality judgment back to the system for search result refinement

# Key components in a search engine

- Search query logs
  - Record users' interaction history with search engine
- User modeling
  - Understand users' longitudinal information need
  - Assess users' satisfaction towards search engine output

# Discussion: Browsing v.s. Querying

- Browsing – what Yahoo did before

- The system informs the user about relevant documents and allows the user to follow the stream of information
- Works well for browsing or does keyword conveyance (e.g., with a smartphone)

- Querying – what Google does

a (keyword) query. The system returns relevant documents when the user enters a query to satisfy her need



# Pull vs. Push in Information Retrieval

- Pull mode – with query
- Push mode – without

The image shows a Google search interface with the query "news about curfew in st louis". The search results page displays approximately 6,140,000 results in 0.34 seconds. The top result is from KMOV.com, titled "1 shot, 7 arrested while police enforced Ferguson curfew", dated 5 days ago. Other results include "Police Begin to Impose Curfew in Ferguson" from CBS Local (6 days ago), "Gov. Nixon declares state of emergency, imposes curfew in ..." from Fox News Channel (7 days ago), "Police begin to impose curfew in St. Louis ..." from Fox News Channel (7 days ago), and "Gov. Nixon taps National Guard to help bring calm in ..." from St. Louis Post-Dispatch (6 days ago). A Yahoo! sidebar is visible on the left, and a weather forecast for Saturday (89°/70°) is on the right.

YAHOO!

Mail

Autos

News

Sports

Finance

Weather

Games

Homes

Health

Beauty

Food

Tech

Answers

Screen

Flickr

Jobs

Shopping

Travel

Dating

More Yahoo Sites >

Google

news about curfew in st louis

Web News Videos Images Shopping More Search tools

About 6,140,000 results (0.34 seconds)

News for news about curfew in st louis

1 shot, 7 arrested while police enforced Ferguson curfew

KMOV.com - 5 days ago

Johnson assured those at the news conference that police would not ... St. Louis County prosecutor Bob McCulloch said it could be weeks ...

Police Begin to Impose Curfew in Ferguson

CBS Local - 6 days ago

More news for news about curfew in st louis

Gov. Nixon declares state of emergency, imposes curfew in ...

www.foxnews.com/.../gov-nixon-declares-emergency-... Fox News Channel

7 days ago - Jay Nixon declared a state of emergency and imposed a curfew Saturday in the St. Louis suburb of Ferguson where black teenager Michael ...

Police begin to impose curfew in St. Louis ... - Fox News

www.foxnews.com/.../police-begin-to-impose-curfew-... Fox News Channel

7 days ago - Police said they fired multiple smoke canisters into a crowd of defiant protesters who gathered in a St. Louis suburb early Sunday where a black ...

Gov. Nixon taps National Guard to help bring calm in ...

www.stltoday.com/news/.../article\_1a915292-8ff... St. Louis Post-Dispatch

6 days ago - ST. LOUIS' #1 SOURCE FOR NEWS · PRINT EDITION · E EDITION · APPS ... Across the street, Aminah Lewis, 36, of St. Louis said she was there to counter ... Highway Patrol: Ferguson curfew remains in effect after midnight.

Sign In Mail

Watch the show »

Toyota cars

Britney Spears

The Sopranos

Justin Gosselin

Justin Cavallari

Protest

and in Mali

Ukraine

the military

fatal shooting

Saturday

89° 70°

98 ↑ 0.05%

11 ↓ -0.04%

80 ↓ -0.02%

# What you should know

- Basic workflow and components in a IR system
- Core concepts in IR
- Browsing v.s. querying
- Pull v.s. push of information