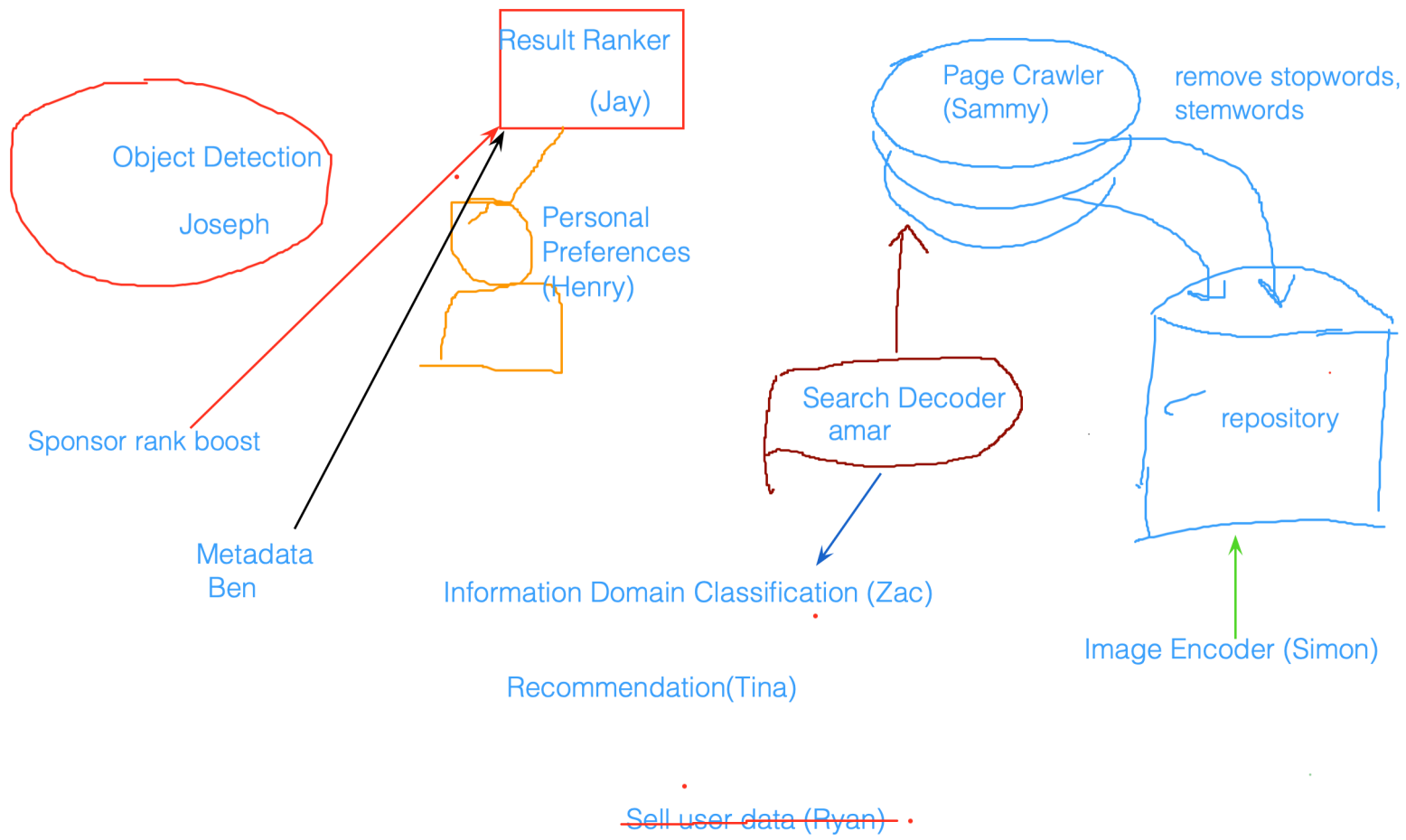


Search Engine Architecture

Hongning Wang

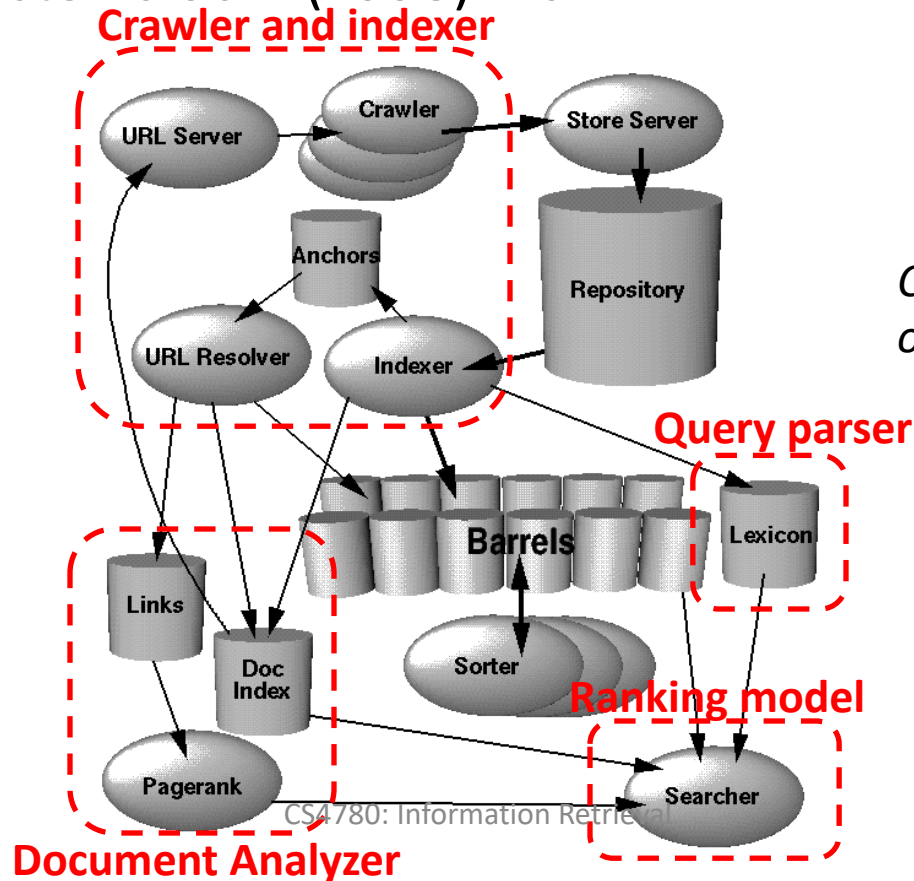
CS@UVa

Crack into Google!



Classical search engine architecture

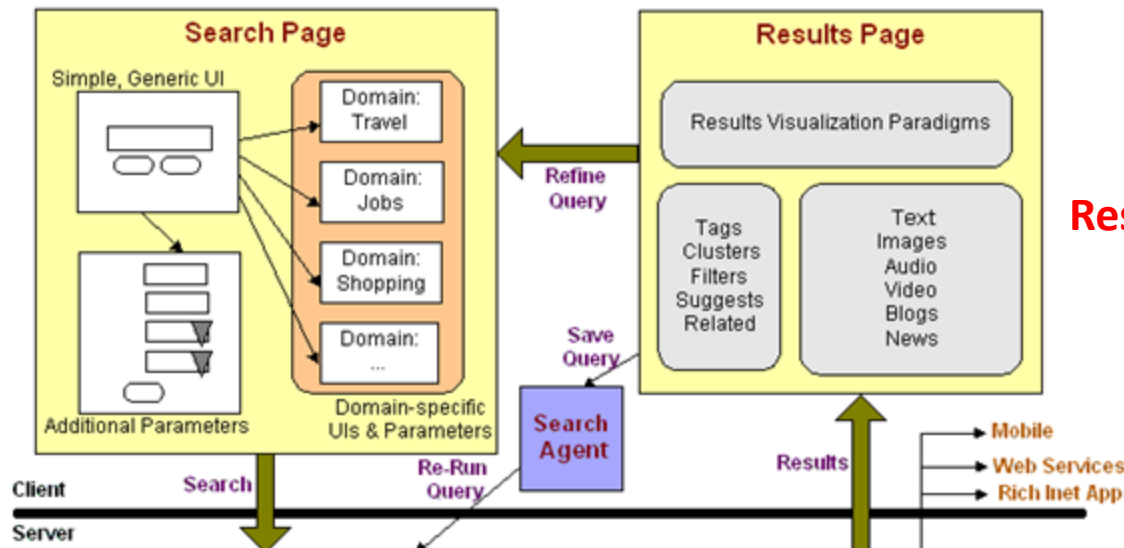
- ***“The Anatomy of a Large-Scale Hypertextual Web Search Engine”*** - Sergey Brin and Lawrence Page, *Computer networks and ISDN systems* 30.1 (1998): 107-117.



Citation count: 20,662 (as of February 11, 2021)

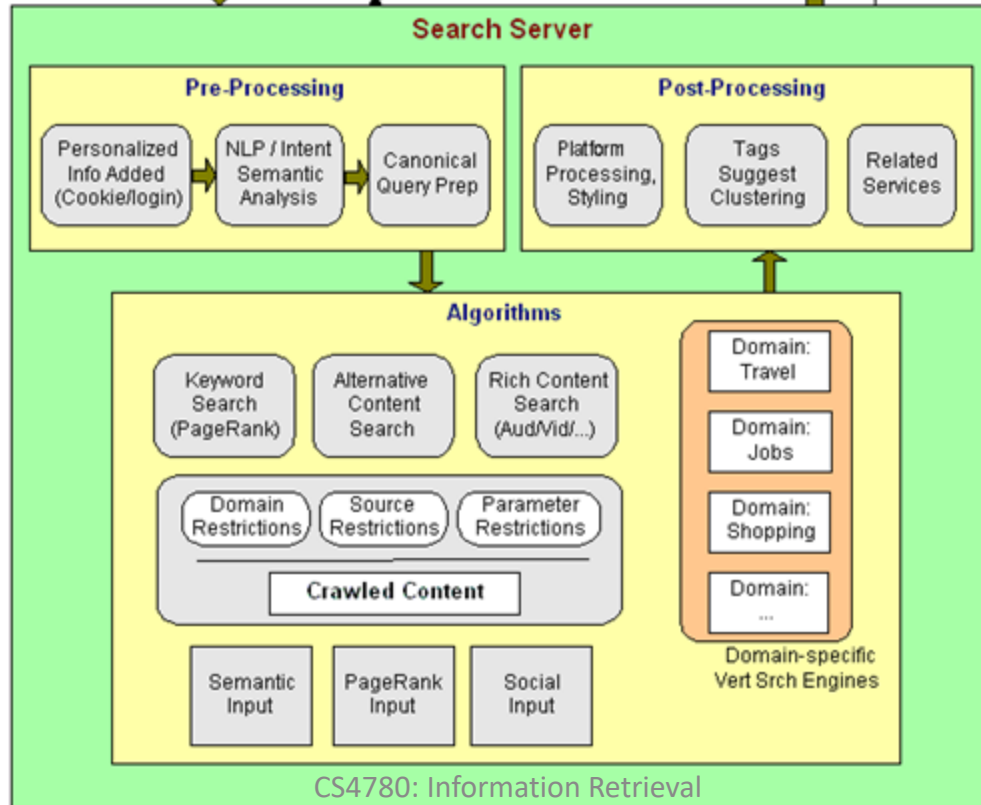
User input

Result display



Query parser

Result post-processing



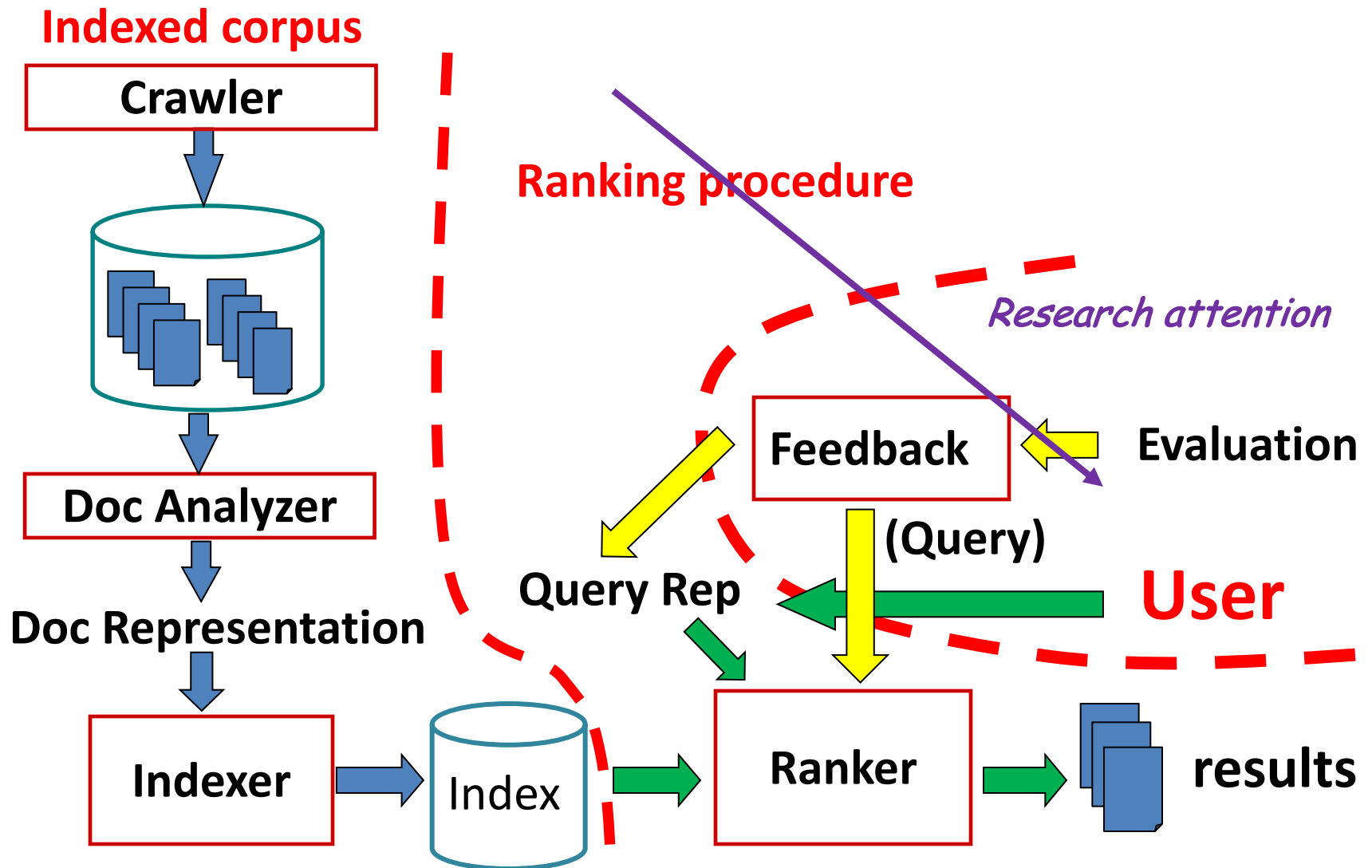
Ranking model

Domain specific databases

Crawler & Indexer

Document analyzer & auxiliary 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***
 - Relate ***the user's information need*** information need
 - Multiple perspectives: topical, semantic, temporal, spatial, and etc.

Key components in a search engine

- Web crawler
 - An automated 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 returned 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 provides information and a relevant document following the structure of the system
- Works well when the user wants to find a document or does a keyword query (e.g., with a smartphone)

- Querying – what Google does now



a (keyword)
e system returns
ant documents
hen the user
y what query to
ssing her
eed

Pull vs. Push in Information Retrieval

- Pull mode – with query
- Push mode – without

The image shows a Google search interface overlaid on a Yahoo! homepage. The search query is "news about curfew in st louis". The search results are displayed in a list format, with the top result being a news article from KMOV.com titled "1 shot, 7 arrested while police enforced Ferguson curfew". Other results include "Police Begin to Impose Curfew in Ferguson" from CBS Local and "Gov. Nixon declares state of emergency, imposes curfew in ..." from Fox News Channel. The Yahoo! homepage in the background shows various navigation links like Mail, Autos, News, Sports, Finance, Weather, Games, Homes, Health, Beauty, Food, Tech, Answers, Screen, Flickr, Jobs, Shopping, Travel, and Dating.

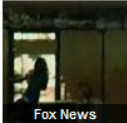
YAHOO!

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

Kate Gosselin

Justin Cavallari

Bank protest

and in Mali

Ukraine

the military

fatal shooting

Saturday
89° 70°

98 ↑ 0.05%

11 ↓ -0.04%

80 ↓ -0.02%

Question: is Yelp a search engine?

The screenshot shows the Yelp homepage for Charlottesville, VA. The search bar contains 'Find pizza' and the location is set to 'Near Charlottesville, VA, United States'. The navigation bar includes categories like Restaurants, Nightlife, Home Services, and options to write a review, attend events, or talk. The main content area displays a list of pizza places with their ratings, prices, and descriptions. A map on the right shows the locations of these businesses in Charlottesville.

Best pizza in Charlottesville, VA Showing 1-10 of 96

Filters: \$, \$\$, \$\$\$, \$\$\$\$ | Open Now | All Filters

- Ad Marco's Pizza** 930 Olympia Dr, Charlottesville, VA 22911 (434) 465-6800. 3 reviews. \$\$\$ - Sandwiches, Pizza, Chicken Wings. We are officially open for business! Try our Ah!lthentic® Italian pizza made with fresh ingredients - dough made in store daily, a 3-cheese blend that is fresh, never frozen, and... [read more](#)
- Ad The Carving Board Cafe** 624 Albemarle Sq, Charlottesville, VA 22901 (434) 974-9004. 40 reviews. \$ - Delis, Caterers, Sandwiches. I just moved here and found this place on Yelp when I needed to satisfy a Reuben craving. Unfortunately, the corned beef had just come out of the oven wasn't totally ready, so I had... [read more](#)
- 1. Lampo** 205 Monticello Rd, Charlottesville, VA 22902 (434) 244-3226. 188 reviews. \$\$\$ - Italian, Pizza
- 2. Uncle Maddio's Pizza** 3912 Lenox Ave, Charlottesville, VA 22901 (434) 234-3717. 11 reviews. \$ - Pizza, Salad. Hard to locate a decent gluten free crust...well Maddios NAILED the delivery!!! Perfect crunch with great sauce to boot!!!! Will be returning! [read more](#)

Mo' Map Redo search when map moved. Ads by Google. [rockefellerfoundation.org](#) Do you agree? - Our core values—are threatened. See curated opportunities from the Rockefeller Foundation to make a difference.

What you should know

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

Today's reading

- Introduction to Information Retrieval
 - Chapter 19: Web search basics