

CS293S Summary

2017 Tao Yang

Search Result Reply Pages

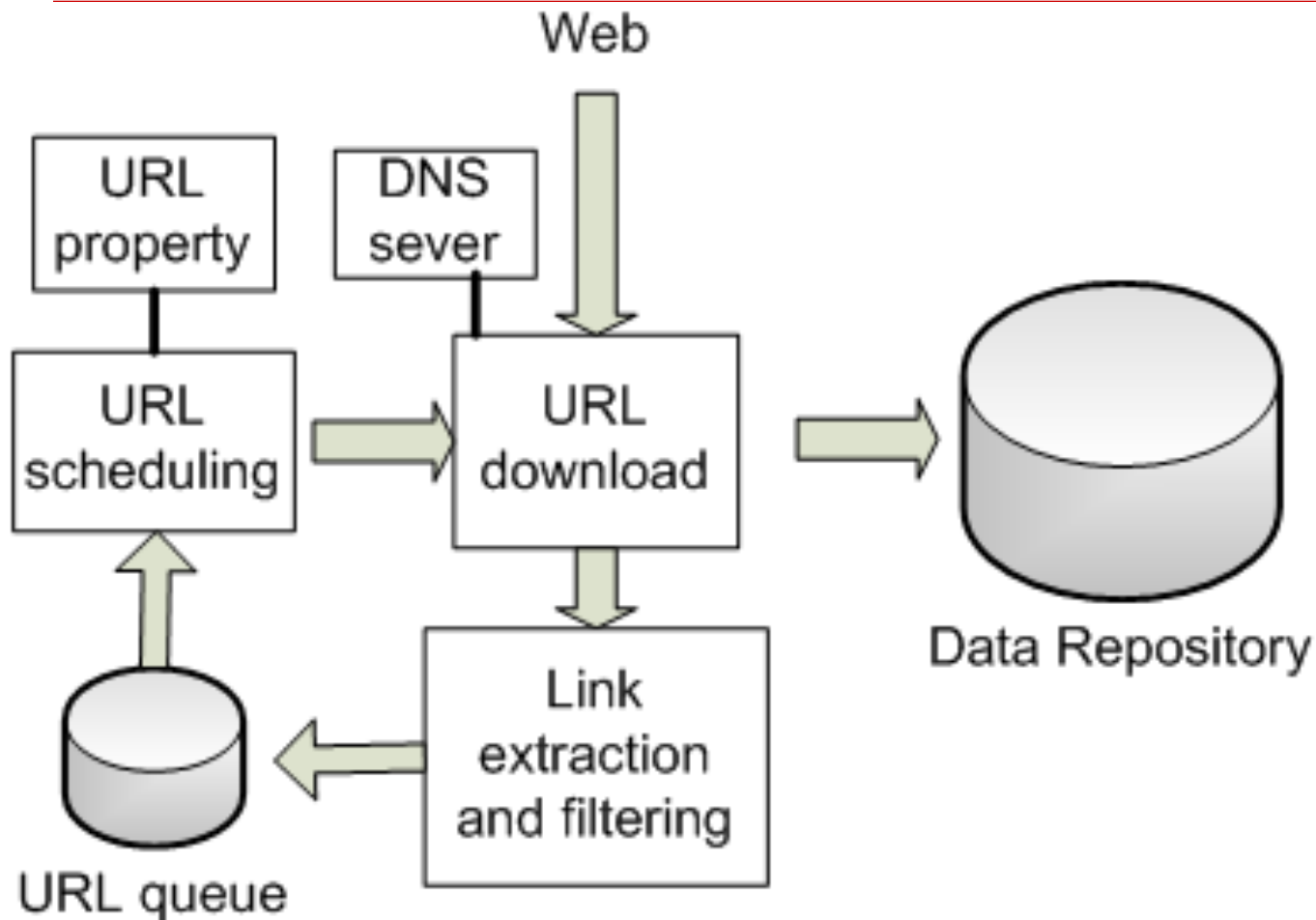
The image shows a screenshot of a web browser displaying a Bing search results page for the query 'ucsb'. The browser's address bar shows the URL 'www.bing.com/search?q=ucsb&go=&q&s=n&form=QBLH&pq=ucsb&sc=8-4&sp=-1&sk='. The search results are categorized into 'Main results' (indicated by a callout), 'Advertisements' (indicated by a callout), and 'Suggestions recommendation' (indicated by a callout). The 'Main results' section includes a link to the official website of the University of California, Santa Barbara, with a description of the university and a list of related links such as 'Admissions', 'Our Campus', 'Academics', 'Administration', 'Working at UCSB', 'Visitors', 'Future Students', and 'Current Students'. The 'Advertisements' section features two ads: 'UCSB Gauchos Store' and 'Ucsb T Shirts'. The 'Suggestions recommendation' section lists related searches like 'UCSB Campus', 'UCSB Sports', 'UCSB Logo', 'UCSB Athletics', 'UCSB Home Page', 'UCSB Directory', 'UCSB Merchandise', and 'UCSB Human Resources'. The browser's taskbar at the bottom shows several open applications, including a PDF viewer and a presentation viewer.

Main results

Advertisements

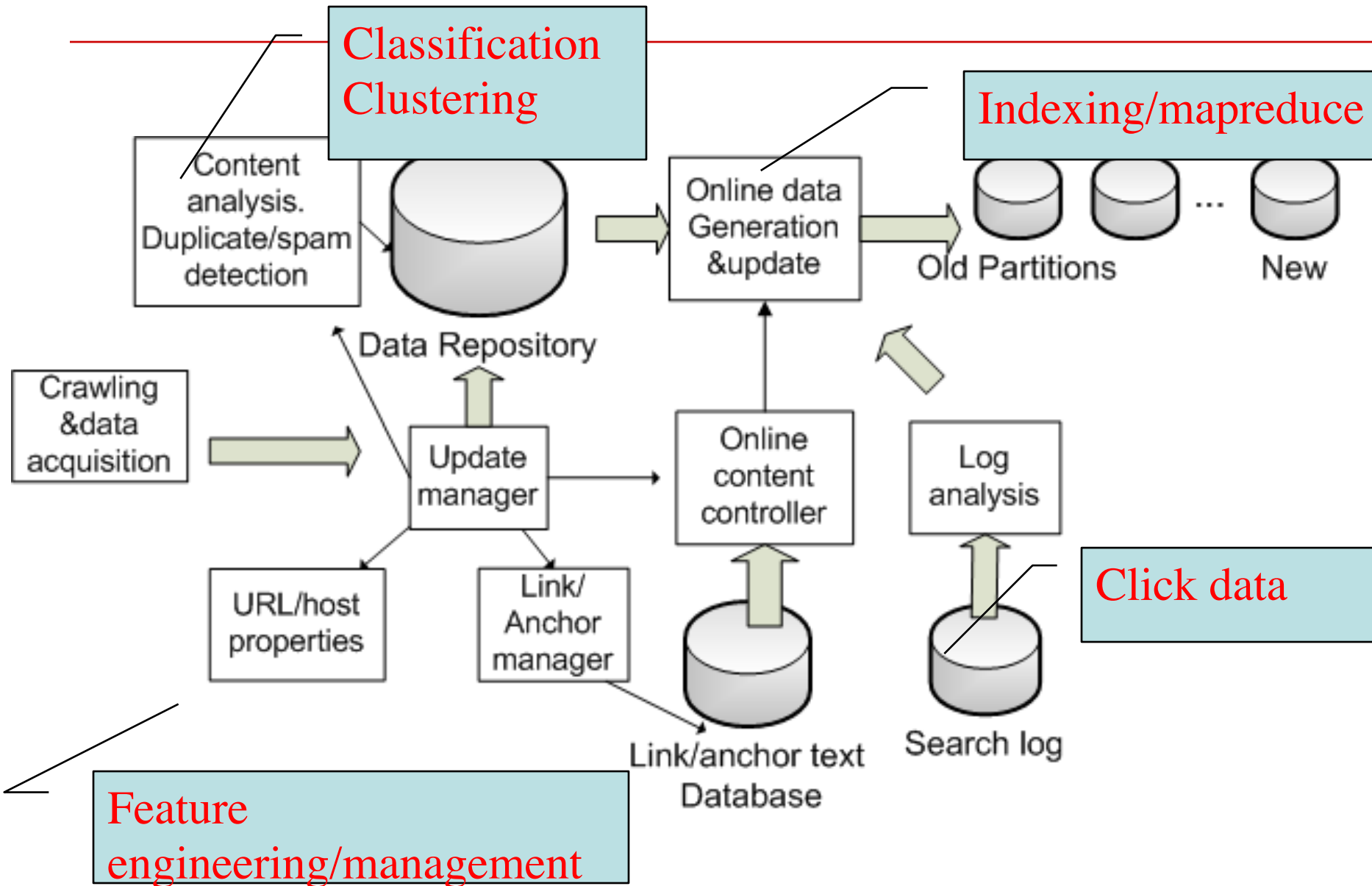
Suggestions recommendation

A Crawler Architecture

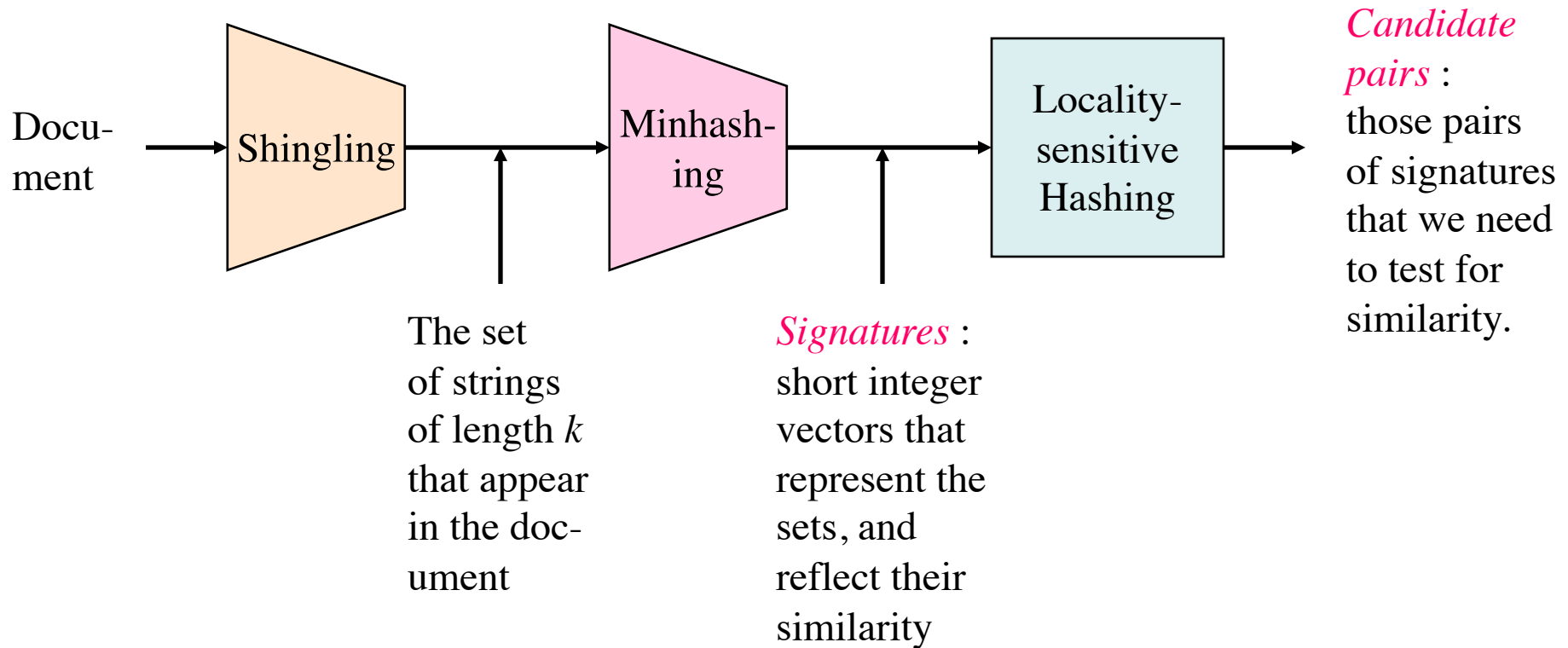


Olston/Najork. Web crawling.
Found. Trends Inf. Retr., 4(3):175--246, March 2010.

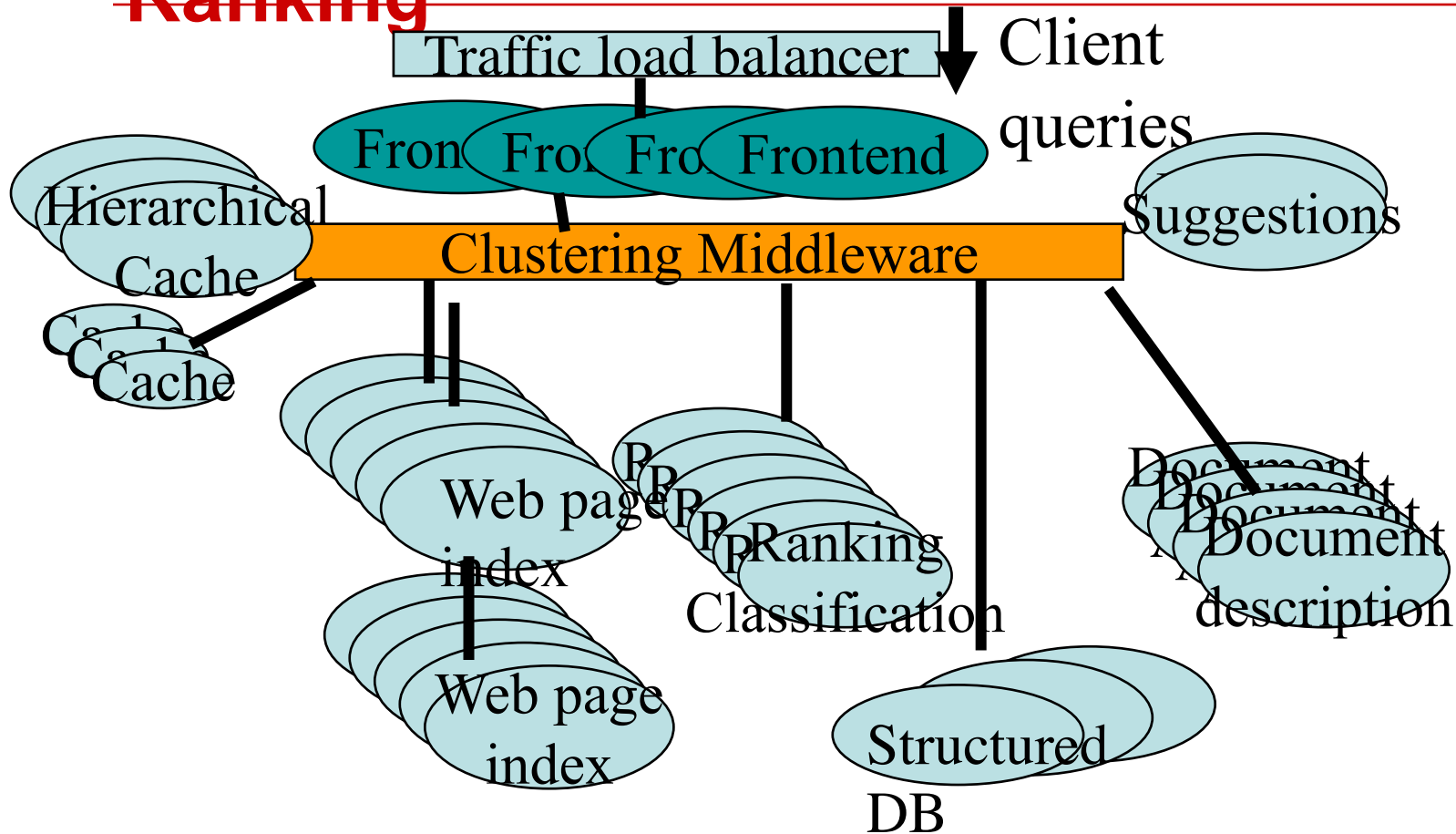
Offline Architecture



Similarity Analysis



Online Engine: Architecture, Matching, Ranking



Web Search for a Planet: The Google Cluster Architecture
[L. Barroso](#), [J. Dean](#), [U. Hölzle](#), IEEE Micro, vol. 23 (2003)

Document Ranking with Text, Quality, and Click Features

- **Text features**
 - TFIDF, BM25
 - Where do they appear? Title/body
 - Proximity (word distance)
- **Document quality and classification**
 - Web link scores (e.g. PageRank).
 - Page length, URL type etc.
- **User behavior data**
 - **Presentation:** what a user sees *before* a click
 - **Clickthrough:** frequency and timing of clicks
 - **Browsing:** what users do *after* a click

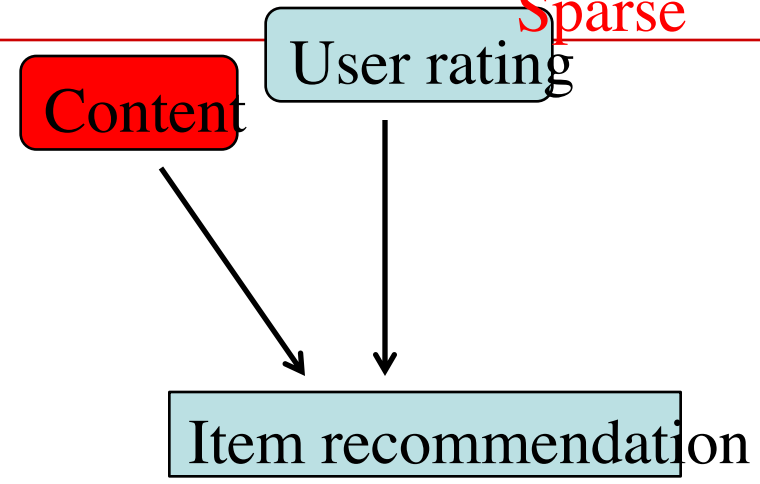
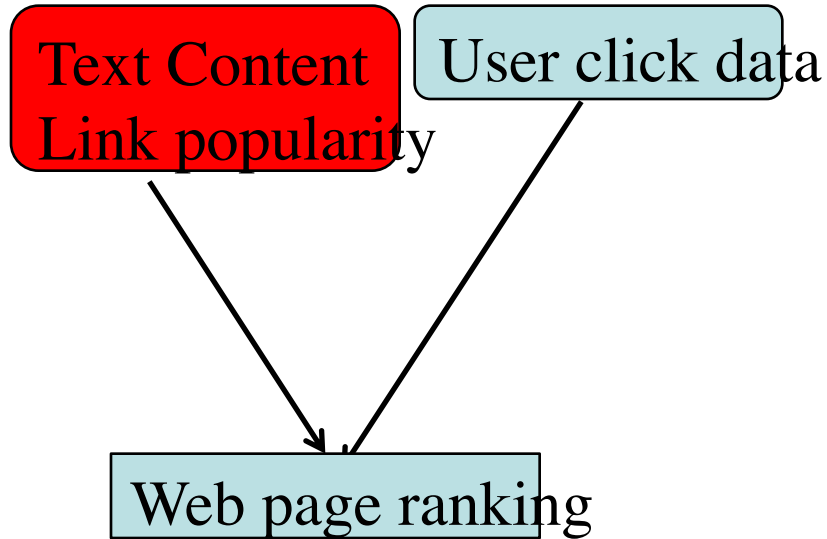
Learning to rank

- Convert ranking problem to a classification problem.
 - *Point-wise* learning
 - Given a query-document pair, predict a score (e.g. relevancy score)
 - *Pair-wise* learning
 - the input is a pair of documents for a query
 - *List-wise learning*
- Bayes, SVM, decision trees, human rules.
- Bagging/boosting to combine multiple schemes

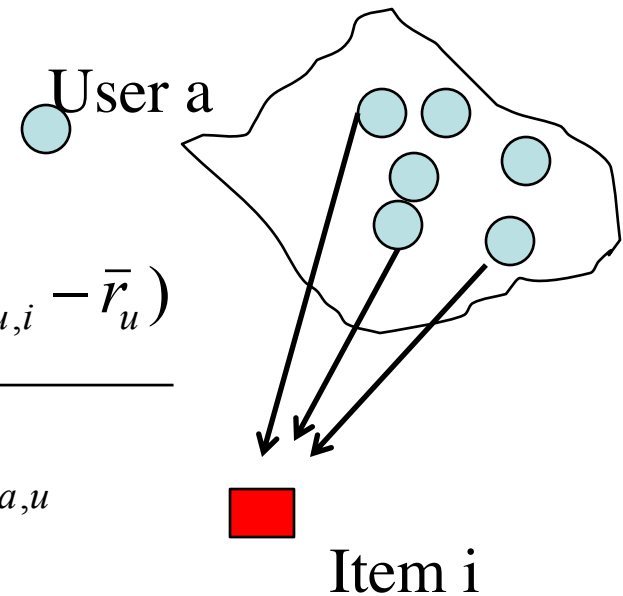
Recommendation vs Search Ranking

Sparse

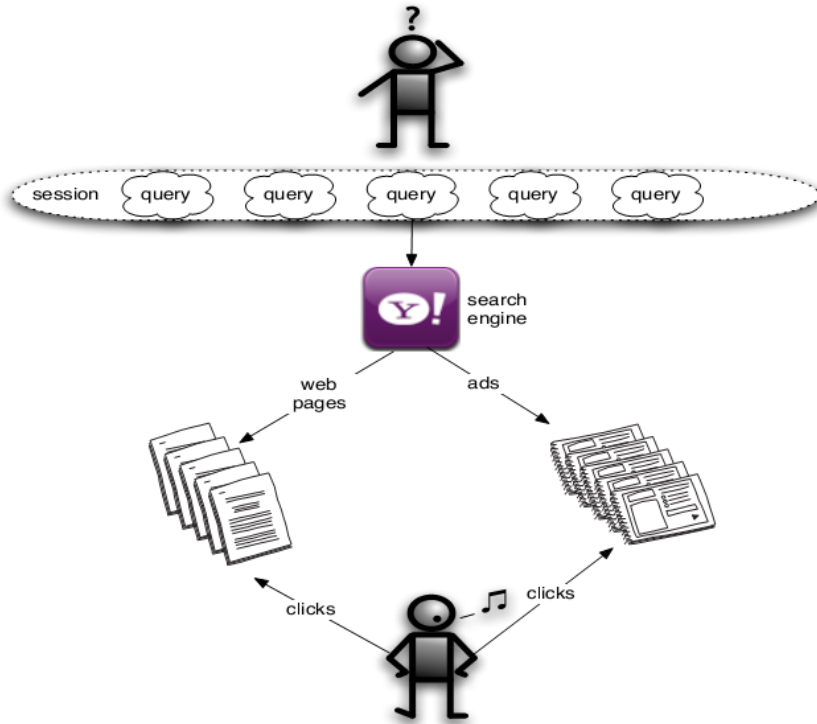
- Collaborative filtering :
Similarity-guided recommendation



$$P_{a,i} = \bar{r}_a + \frac{\sum_{u=1}^n w_{a,u} (r_{u,i} - \bar{r}_u)}{\sum_{u=1}^n w_{a,u}}$$



Search Advertisement



Bid phrase: computational advertising
Bid: \$0.5

Title { [ACL-08:HLT Tutorial](#)

Creative { Computational Advertising Tutorial
Columbus, OH - June 15, 2008

Display URL { [research.yahoo.com](#)

Landing URL: http://research.yahoo.com/tutorials/acl08_compadv/