

Code Review: E-commerce Ads Search System

by
Yang Qiao
Haoyu Xu

User Stories

By calculating the relativity between the user's query and the ads provided by merchants, this ads searching system displays the most relevant ads that users are most interested in so that user are more likely to click the ads.

In the meantime, merchants can easily target the most potential customers and put the right ads with the lowest cost.

User Stories

Advertiser:

- Create Ads associated with bid price
- Bid for position
- Pay by CPC (cost per click)

User:

- Send query to search engine, expressing some intent

Search Engine:

- execute query against web corpus
- execute query against ads corpus
- Display search result page : web result, ads

System Schema

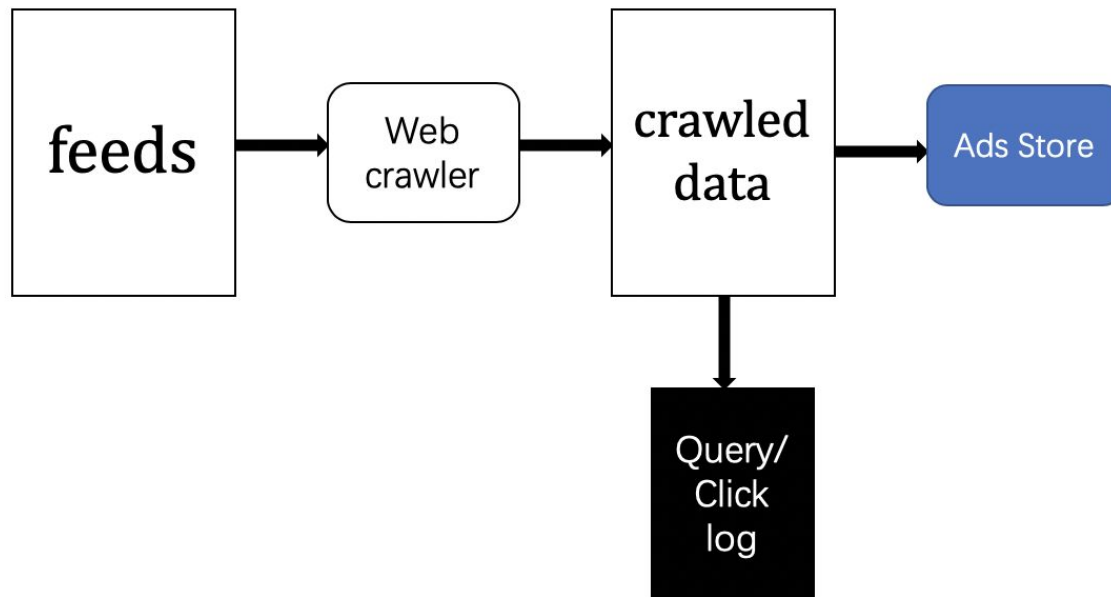
- Query processing :
 - a. Cleaning unnecessary word
 - b. Intent Prediction
 - c. Topic expansion:nike running shoe -> nike sneakers
- Selecting ads :
 - a. apply information retrieval algorithm on index server
- Ranking ads by certain rules:
 - a. relevance between query and ad
 - b. click probability
 - c. bid price
- Select Top K ads by rankings

Web Crawler

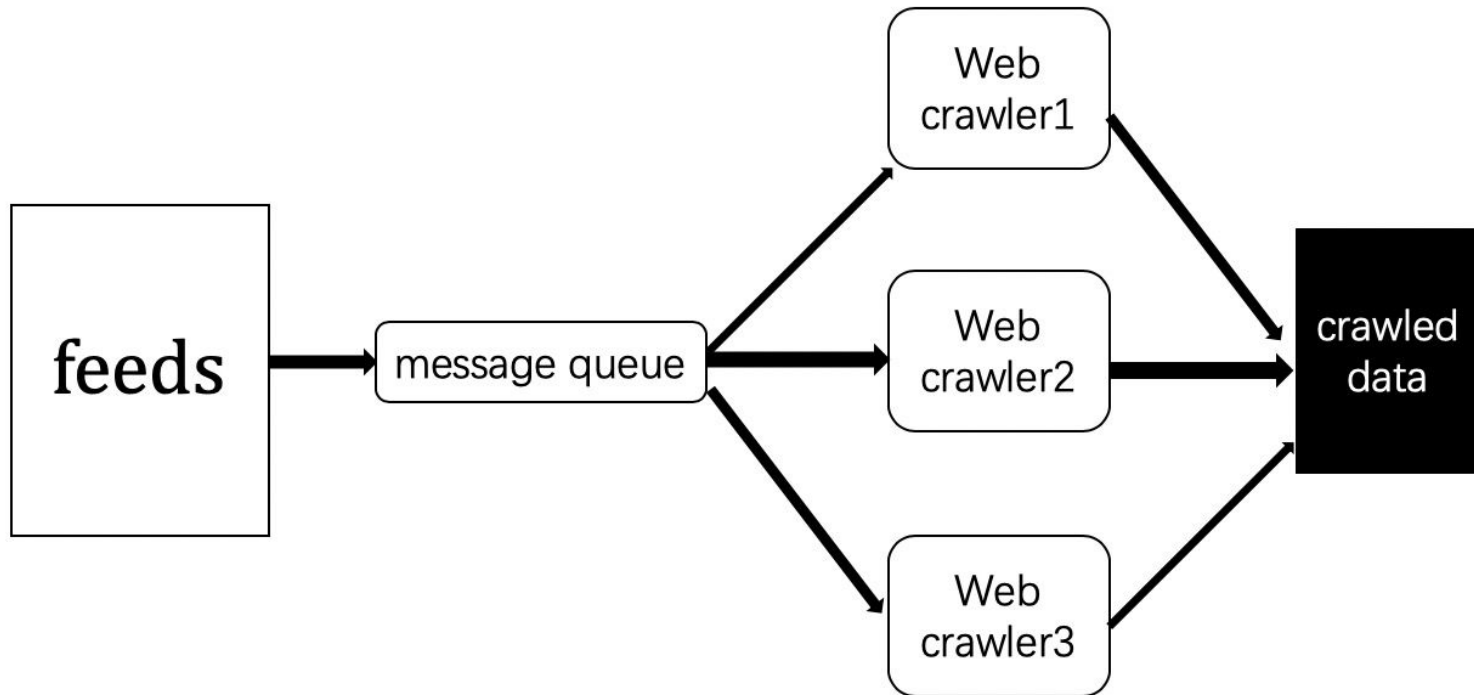
They have two choice for Making requests and extracting data from the response:

1. Parse HTML page synchronously
2. Parse HTML page asynchronously

Web Crawler: Synchronously



Web Crawler: Asynchronously



Comparison: Sync & Async

- Synchronously
 - Pros: Save disk space
 - Cons:
 1. If need extract more data then have to crawl again
 2. Crawling is blocked by parsing HTML
- Asynchronously
 - Pros:
 1. Re-parse HTML if needed
 2. unblock crawling from parsing HTML
 - Cons:
 1. Need more storage space
 2. Need more machines or CPU to parse HTML

Prediction: Search Log

- Data format: Device IP, Device id, Session id, Query, AdId, Clicked(0/1), Ad category_X_Query category_matched
- Goal: generate click log for query intent prediction and click prediction
- Method: reverse engineering
- Positive feature: IP, device_id, AdId, QueryCategory_AdsCategory match, Query_CampaignLd match, Query_AdId match
- Negative feature: mismatached query_category ads_category, mismatched Query_CampaignId, mismatched Query_AdId, lowest campaignId weight, lowese add weight per query group