Related Searches at LinkedIn

Mitul Tiwari
Joint work with Azarias Reda, Yubin Park, Christian Posse, and Sam Shah
LinkedIn

Who am I

Mitul Tiwari

Senior Research Engineer; Search, Network and Analytics, LinkedIn

San Francisco Bay Area | Computer Software

Previous	Kosmix, Google Inc., Microsoft Corp.
Education	PhD, Computer Science at The University of Texas at Austin

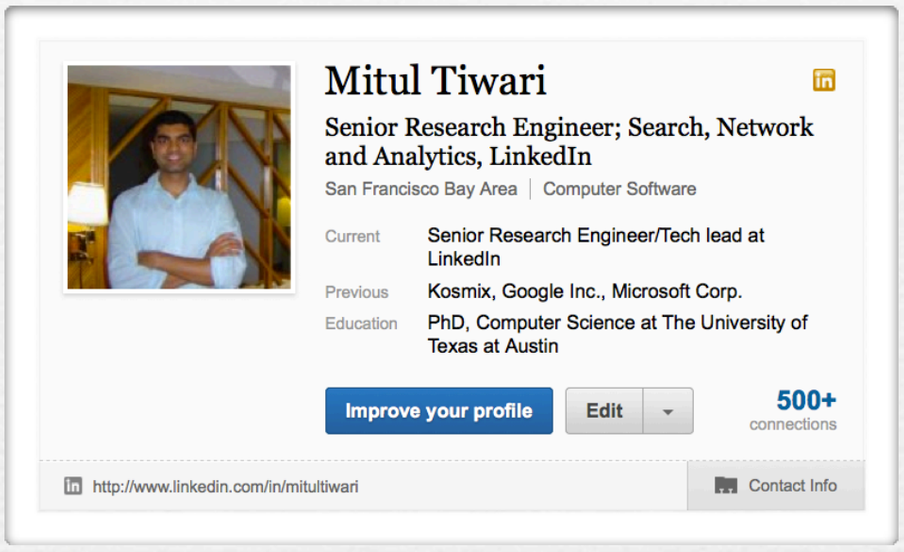
Outline

- About LinkedIn
- Related Searches
  - Design
  - Implementation
  - Evaluation

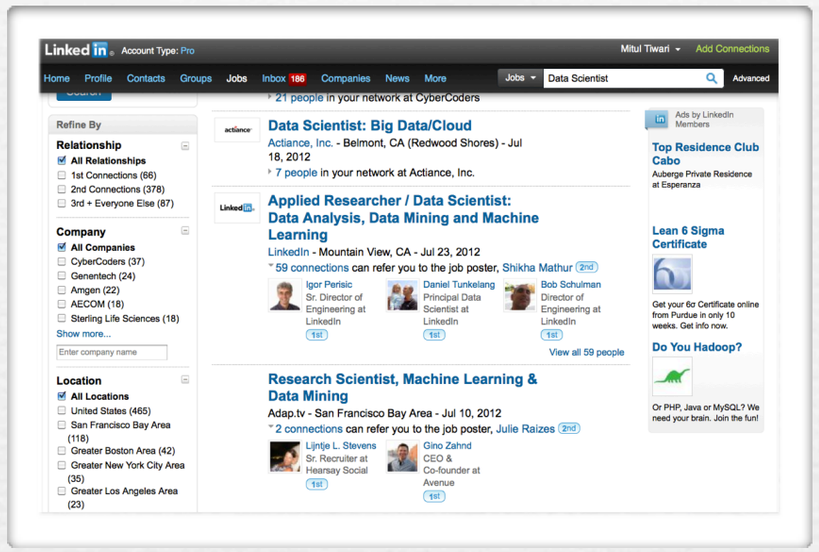
LinkedIn by the numbers

- 175M+ members
- 2+ new user registrations per second
- 4.2 Billion people searches in 2011
- 9.3 Billion page views in Q2 2012
- 100+ million monthly active users in Q2 2012

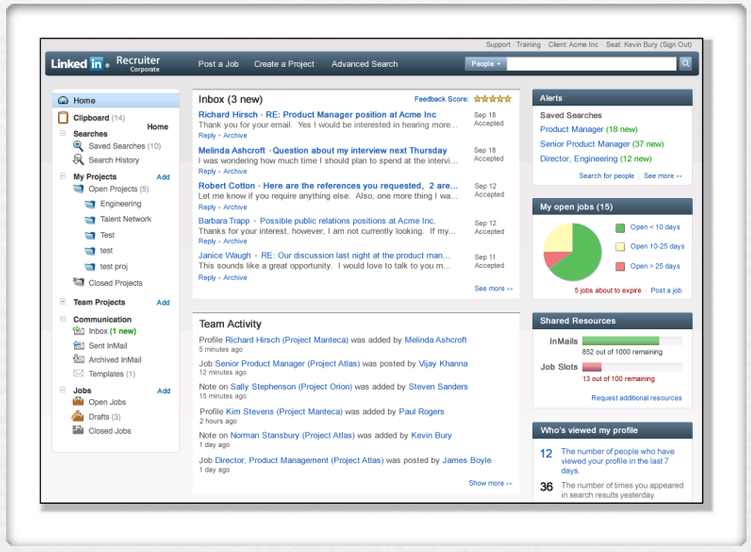
Broad Range of Products



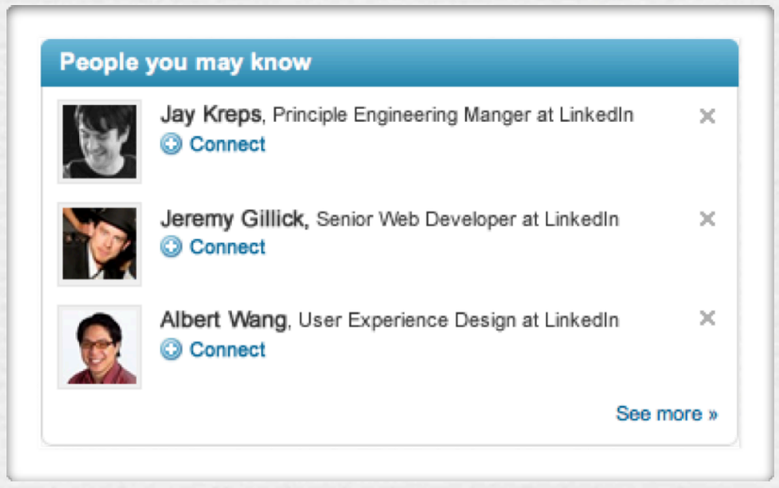
Profile



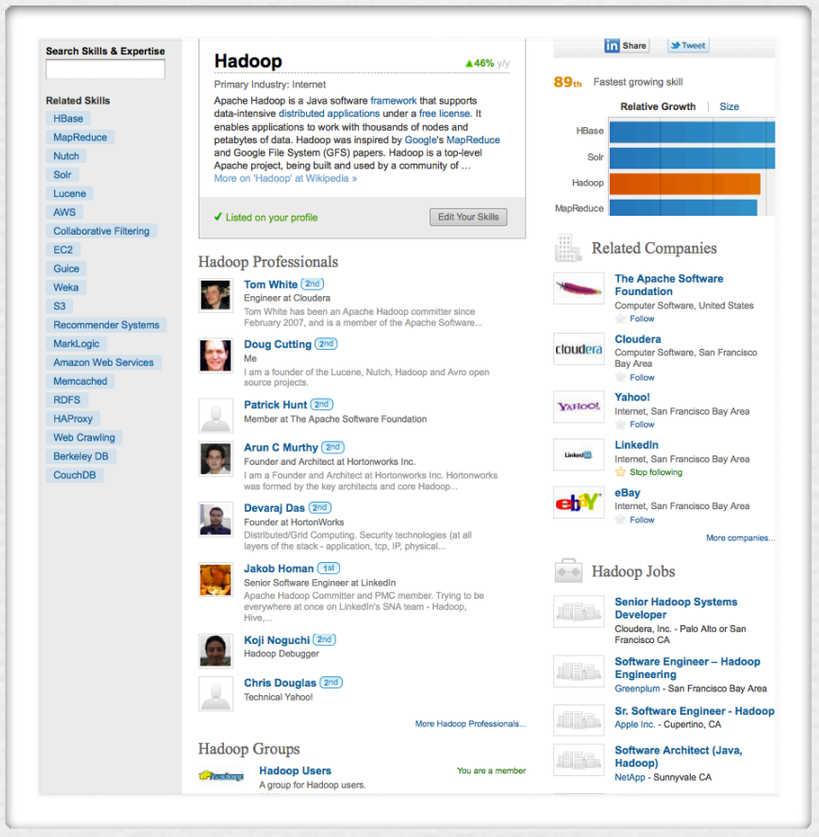
Search



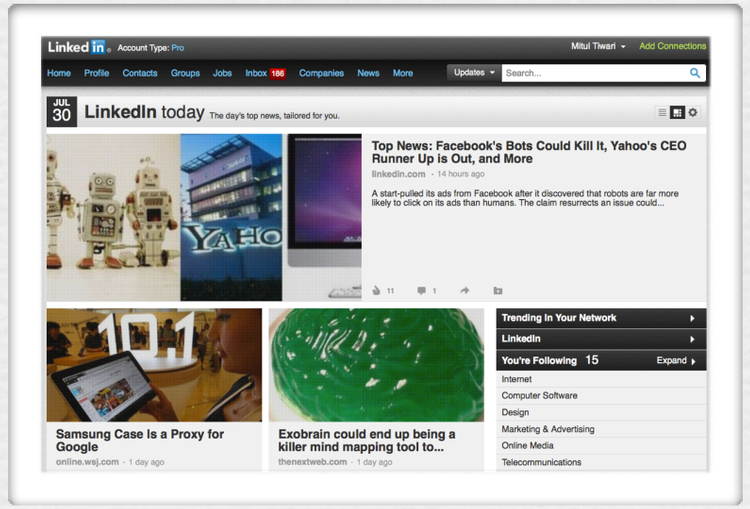
Hiring Solutions



People You May Know



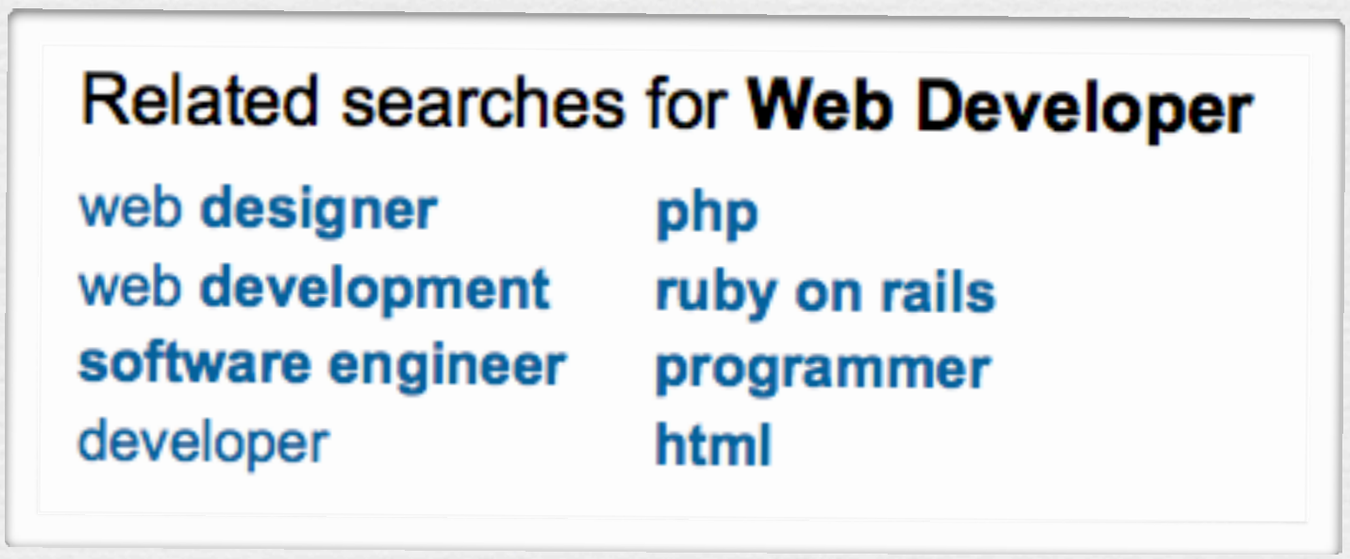
Skills



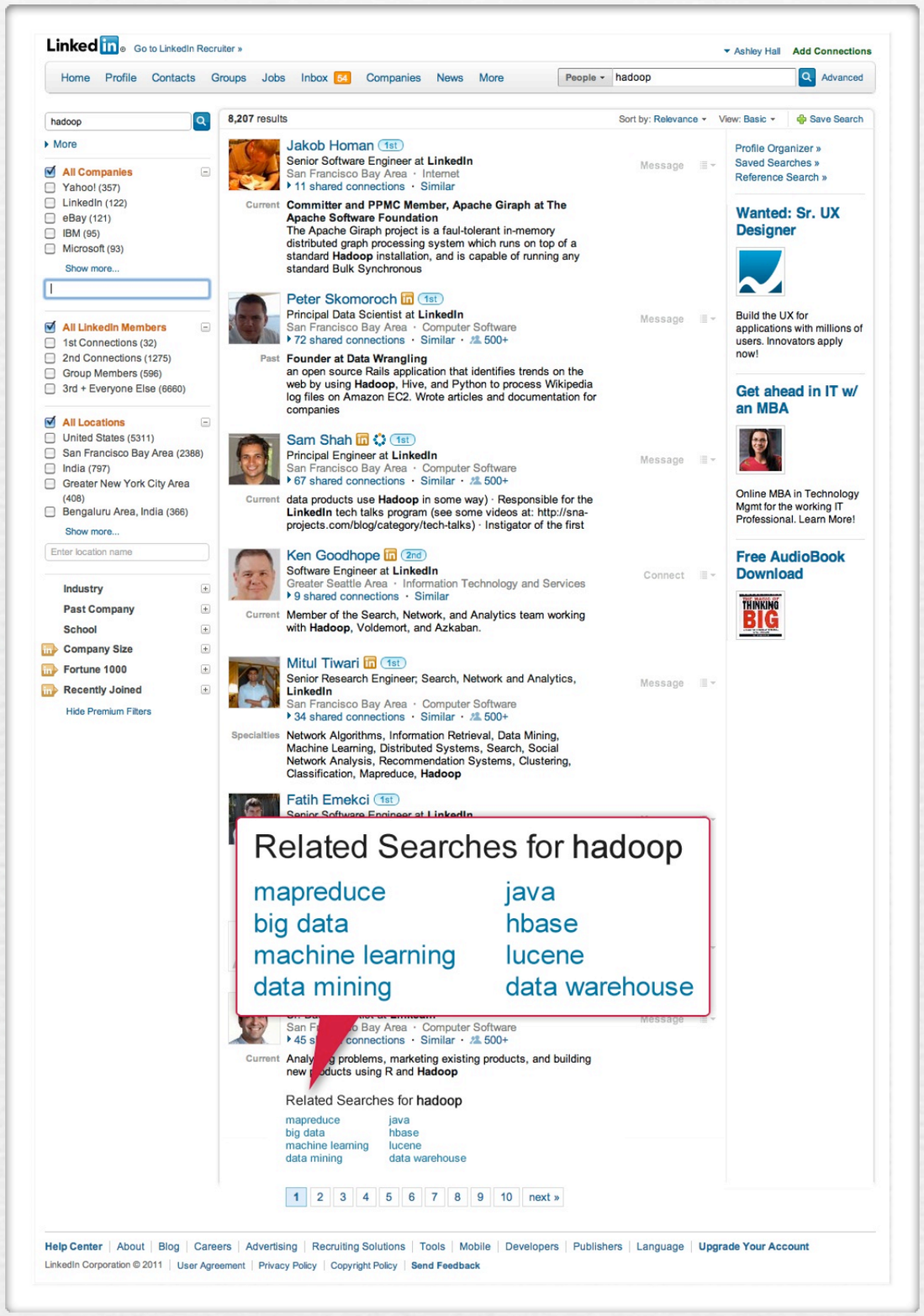
News

Related Searches at LinkedIn

- Millions of searches everyday
- Goal: Build related searches system at LinkedIn
- To help users to explore and refine their queries



Related Searches at LinkedIn



Related Searches

- Design
- Implementation
- Evaluation

Related Searches

- Design
- Implementation
- Evaluation

Design

Design

- Signals
  - Collaborative Filtering
  - Query-Result Click graph
  - Overlapping terms

Design

- Signals
  - Collaborative Filtering
  - Query-Result Click graph
    - Overlapping terms
- Length-bias

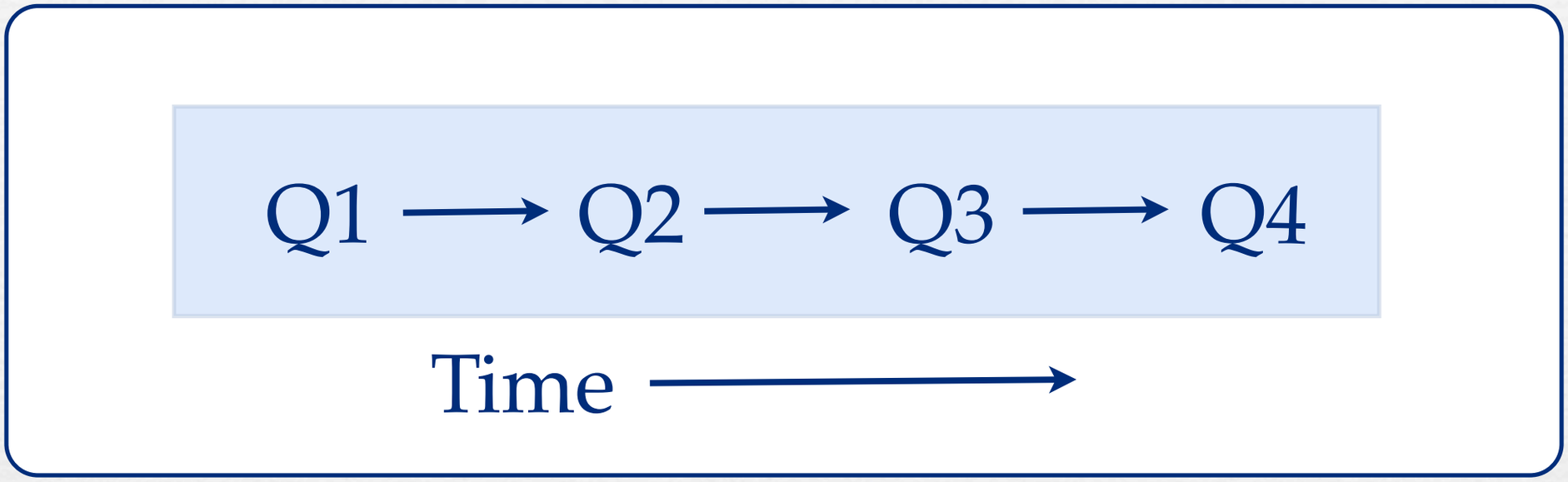
Design

- Signals
  - Collaborative Filtering
  - Query-Result Click graph
    - Overlapping terms
- Length-bias
- Ensemble approach for unified recommendation

Design

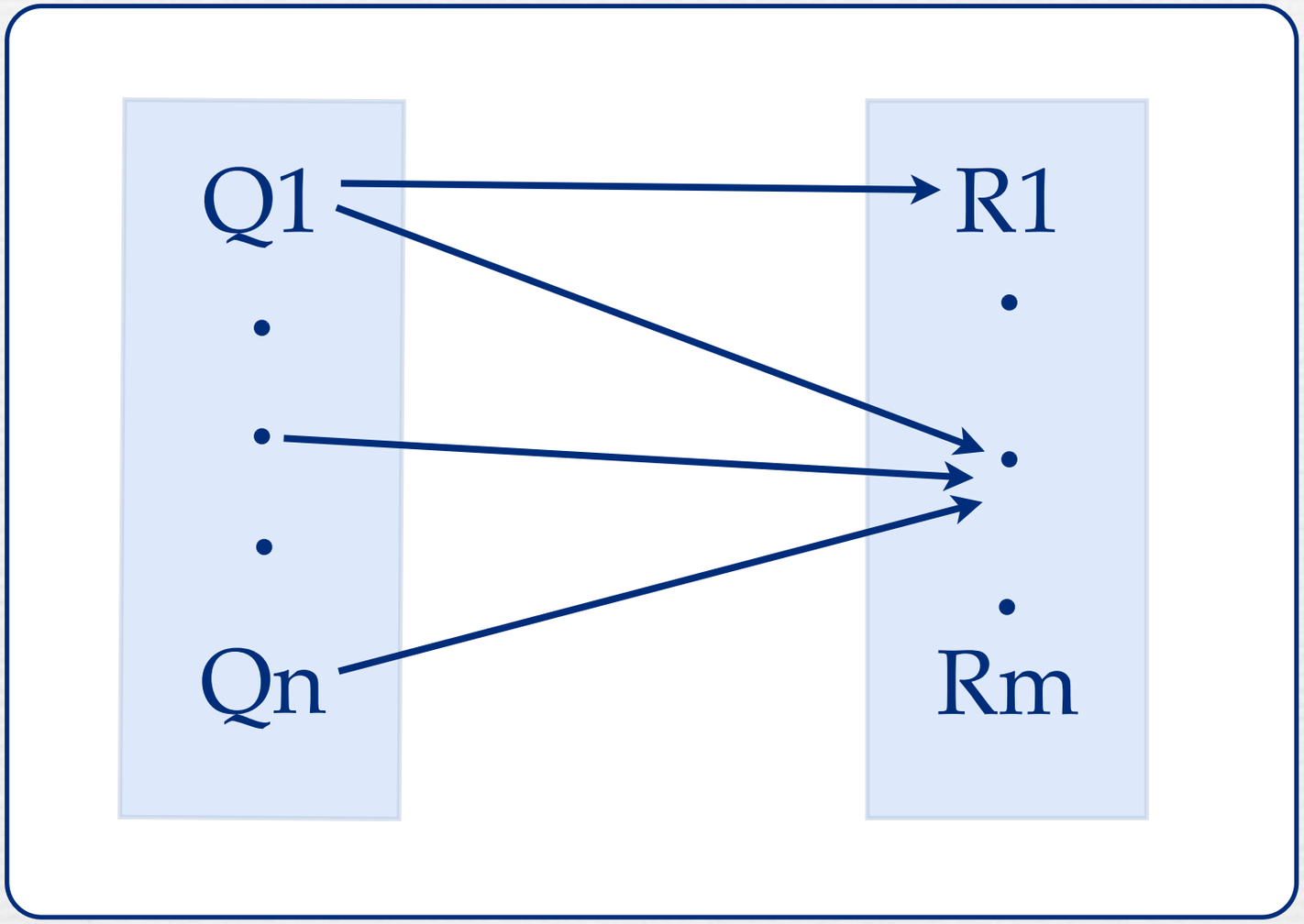
- Signals
  - Collaborative Filtering
  - Query-Result Click graph
    - Overlapping terms
- Length-bias
- Ensemble approach for unified recommendation
- Practical considerations

Design: Collaborative Filtering



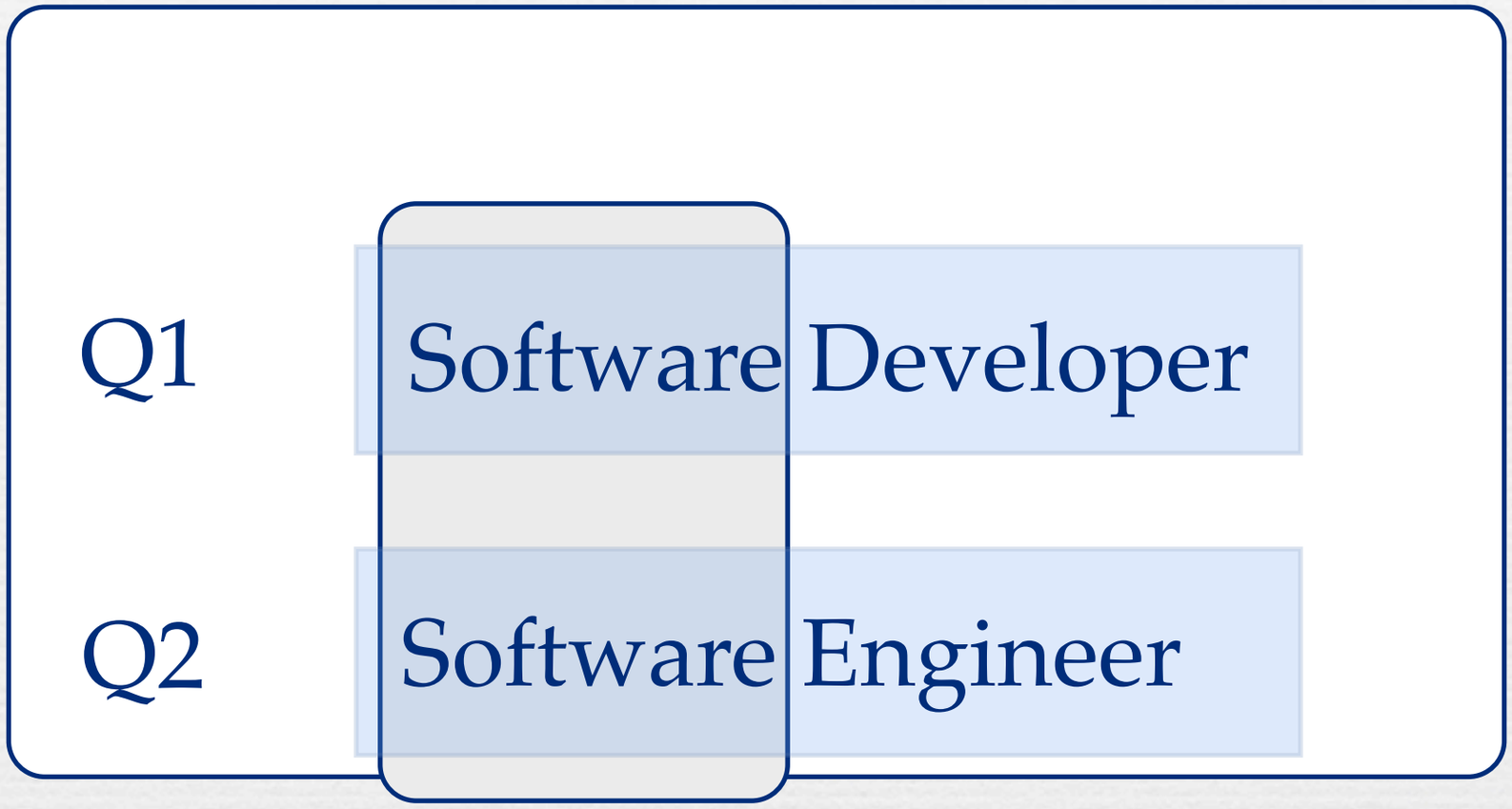
- Searches correlated by time
  - Searches done in the same session by the same user
  - Collaborative filtering: implicit feedback
  - TFIDF scoring to take care of popular queries (e.g. `Obama')

Design: Query-Result Clicks



- Searches correlated by result clicks

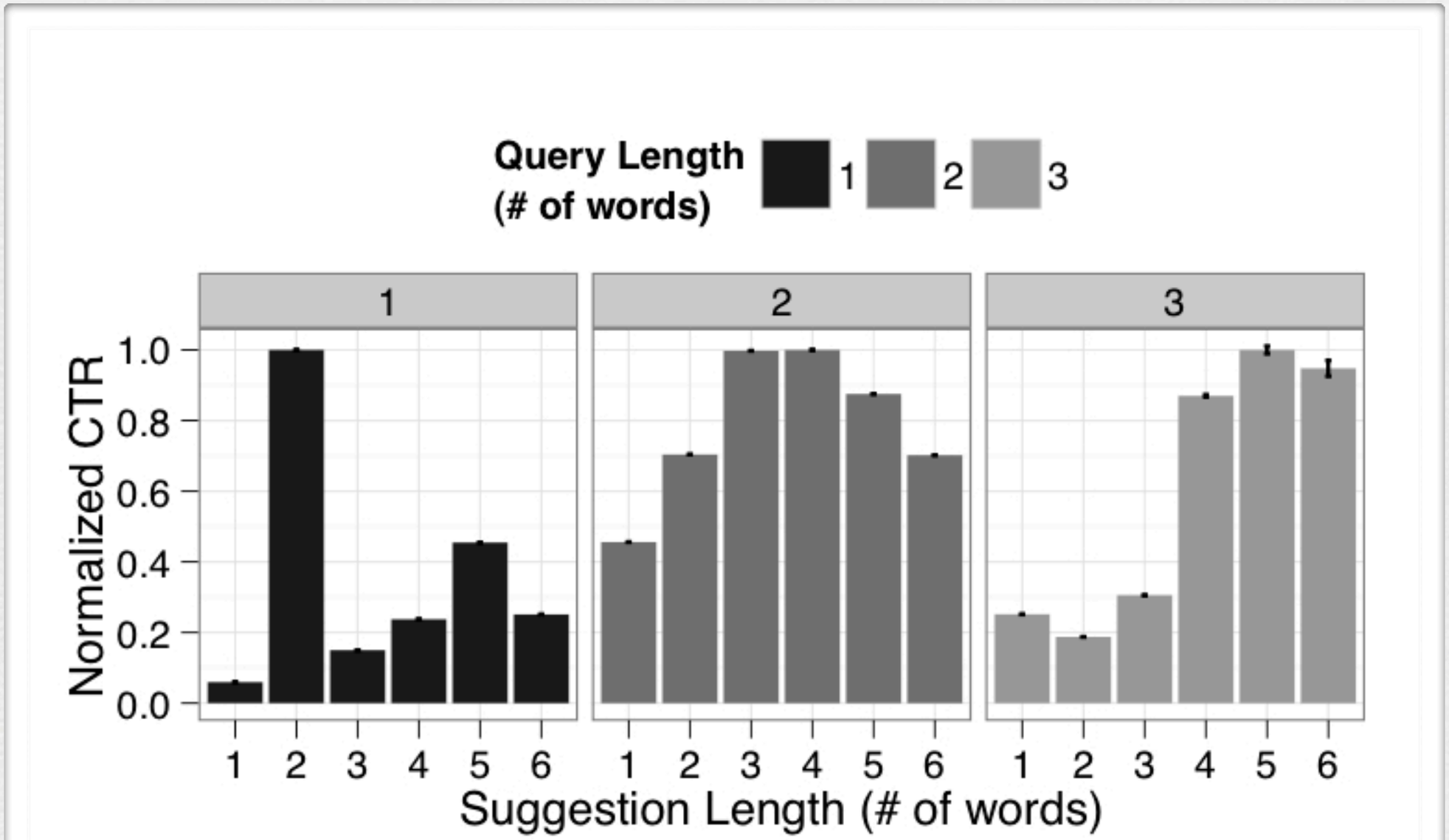
Design: Overlapping Terms



- Searches with overlapping terms
  - TFIDF scoring to give importance to terms

Design: Length Bias

- Insight: clicks on suggestions one term longer



Design: Length Bias

- Insight: clicks on suggestions one term longer
- Corresponds to refining the initial query
- Statistical biasing model to score a longer query higher

Design: Length Bias

- Insight: clicks on suggestions one term longer
- Corresponds to refining the initial query
- Statistical biasing model to score a longer query higher

$$S' \leftarrow S + \lambda \cdot \delta$$

Design: Ensemble Approach

- Need to generate unified recommendation dataset
- Analysis to figure out engagement of each signal
- Attempted ML approach
  - Minimal overlap across different signals

Design: Ensemble Approach

- Step-wise unionization
- Importance based on individual signal performance
  - First, collaborative filter
  - Second, queries correlated by query-result clicks
  - Third, queries overlapping terms

Design: Practical Considerations

- System designed for public consumption
  - Strong profanity filters
  - Need to deal with misspellings
  - Languages
  - Remove spammy search queries

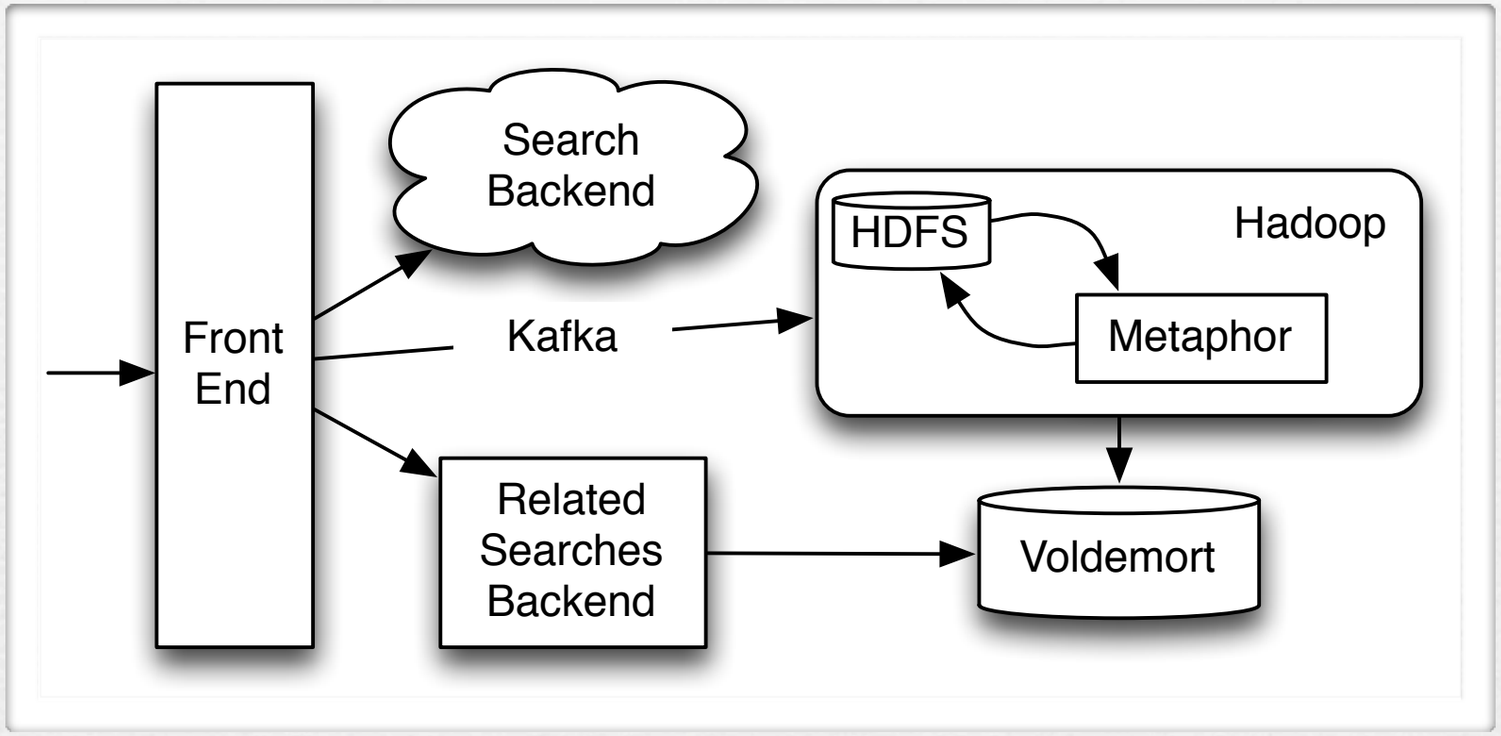
Related Searches

- Design
- Implementation
- Evaluation

Implementation Challenge

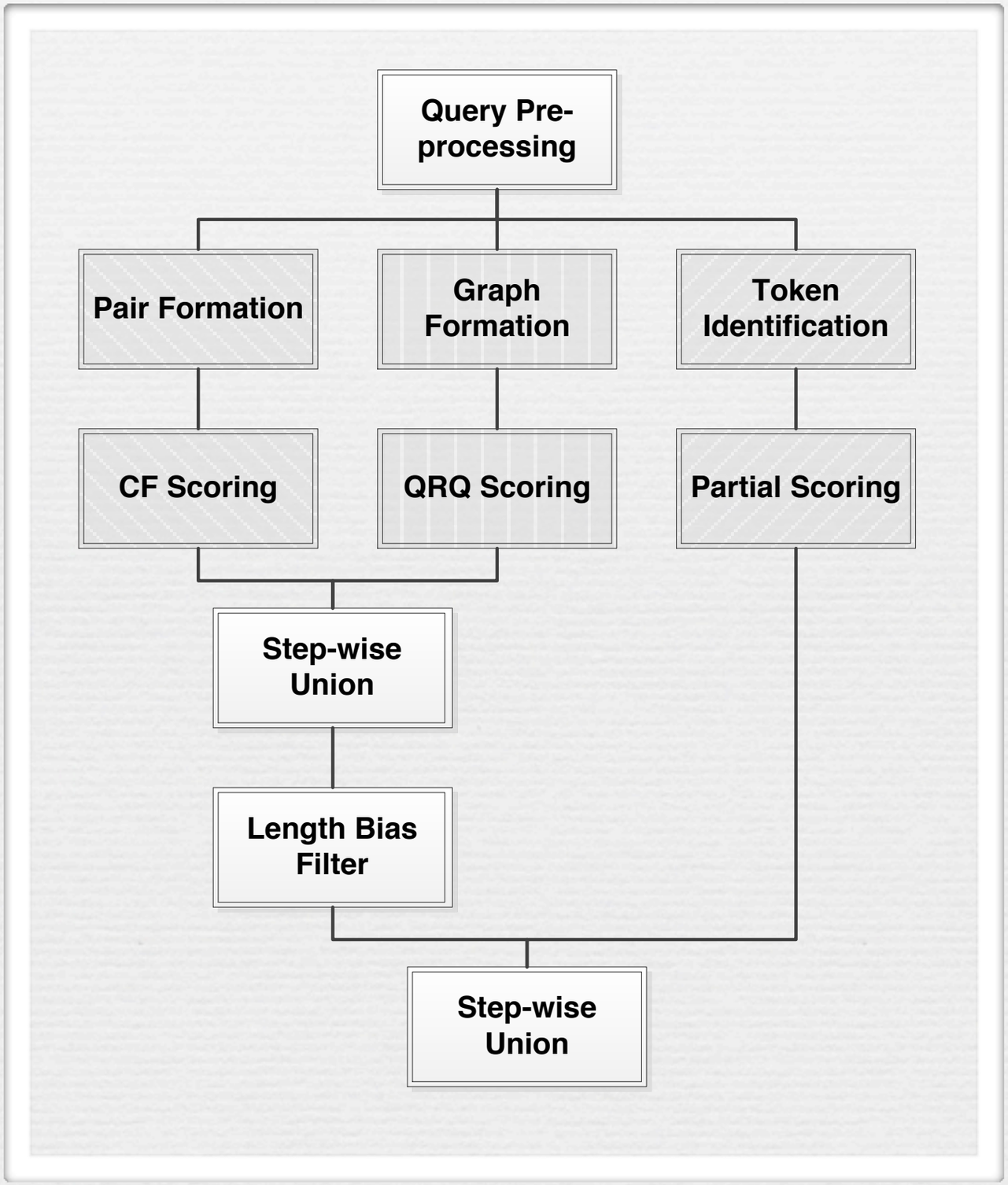
- Scale
  - 175M+ members
  - Billions of searches
  - Terabytes of data to process

Implementation



- Kafka: publish-subscribe messaging system
- Hadoop: MapReduce data processing system
- Azkaban: Hadoop workflow management tool
- Voldemort: Key-value store

Implementation: Workflow



Related Searches

- Design
- Implementation
- Evaluation

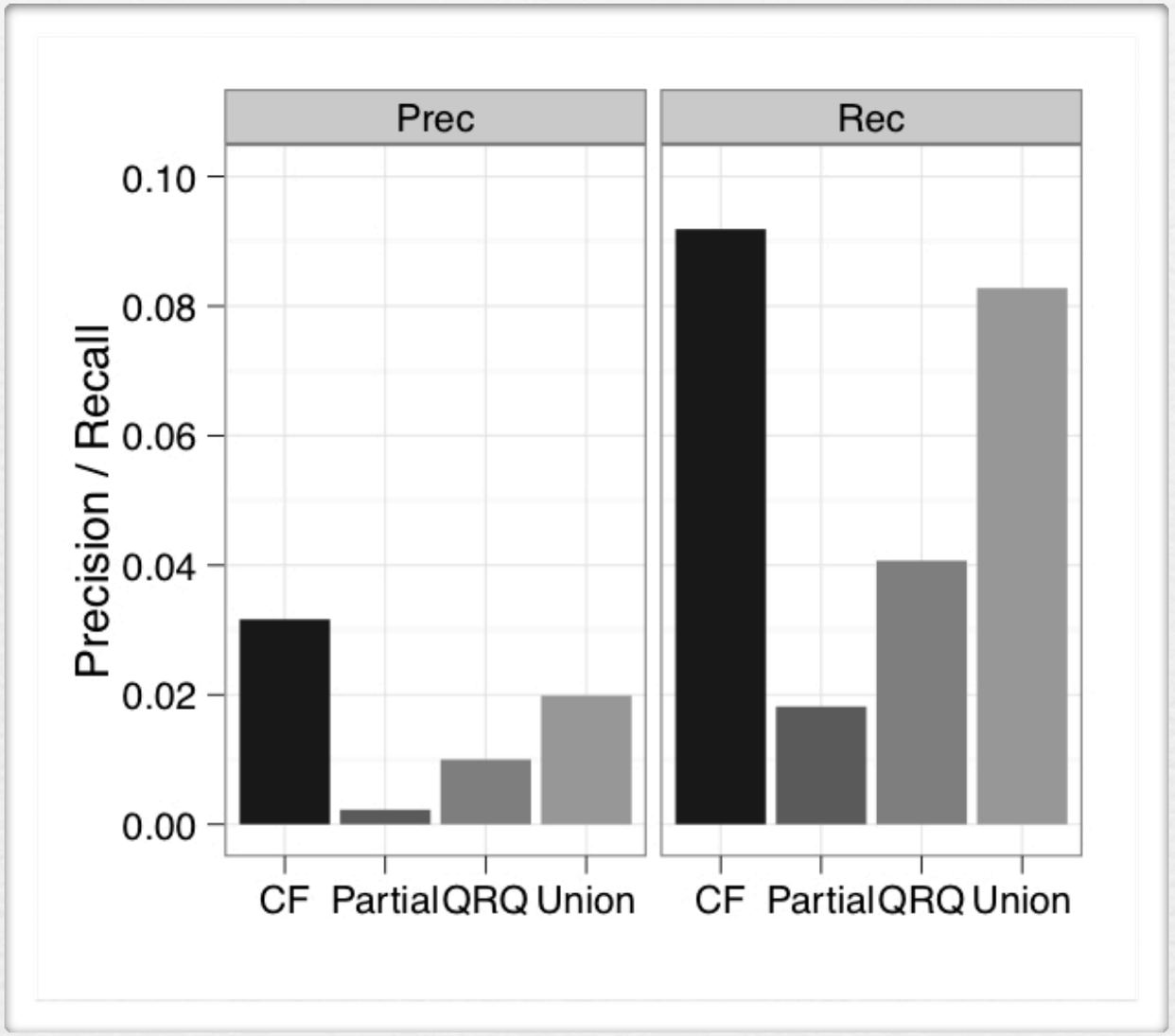
Evaluation

- Performance of each signal and combination
- How does the system scale?

Evaluation Cont'd

- Offline evaluation
  - Precision-Recall
- Online evaluation
  - A/B testing to measure engagement
  - Performance evaluation

Offline Evaluation

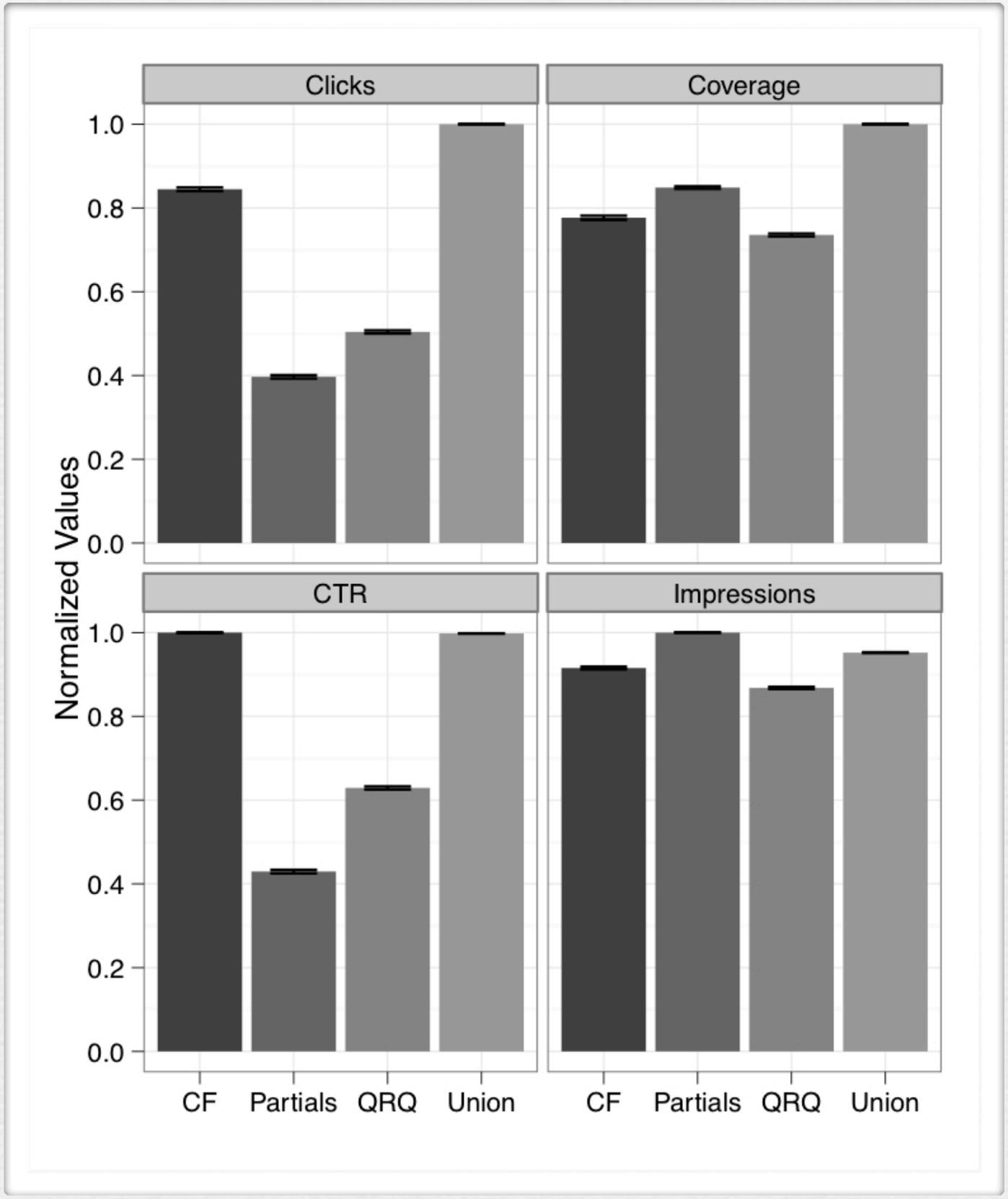


- Correct set: set of searches performed by a user in the following K minutes, here K=10

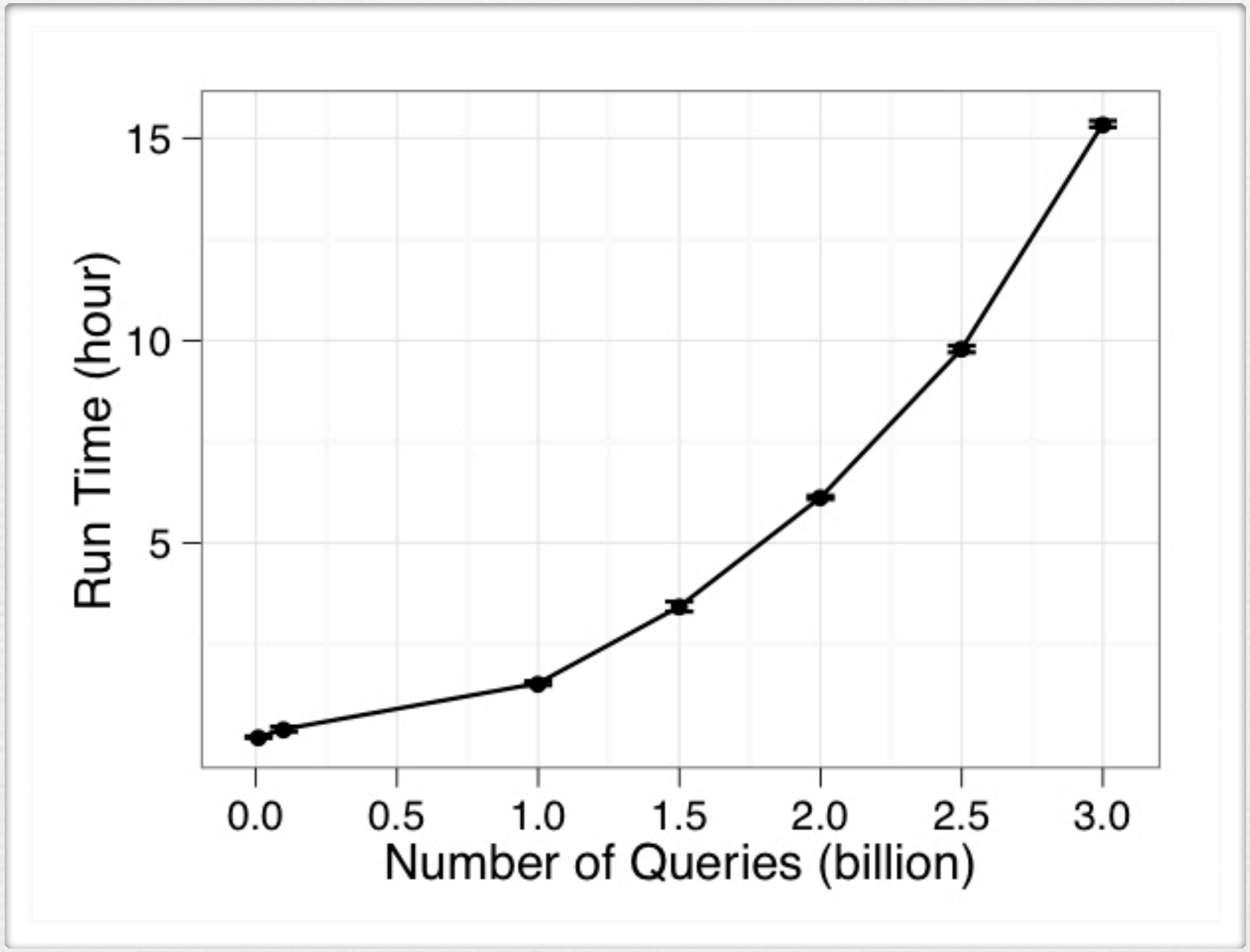
Online Evaluation

- Used A/B testing
- Metrics
  - Coverage: queries with recommendations
  - Impressions: # of recommendations shown
  - Clicks: Clicks on recommendations
  - Click-through rate (CTR): Clicks per impression

Online Evaluation



Evaluation: System Runtime



Details

- Metaphor: a System for Related Search Recommendations, Azarias Reda, Yubin Park, Mitul Tiwari, Christian Posse, and Sam Shah. In *Proceedings of the CIKM*, 2012.