ColumbuScout
Towards Building Local Search Engines over Large Databases

By: Cody Hansen
Professor: Feifei Li

Small Site Search
- Why have a form based search?
  - Slow
  - No Spell Checking
  - No Suggestions
- Search in Specific Boxes

Large Site Search
- When we are all used to this.
  - Fast
  - Spell Checking
  - Suggestions
  - Search Any Combination of Words

A Search Engine With the Functionality of a Large Search Engine that is Deployable on your own Database

Uses:
- Ranking Based on Rarity and Accuracy
- Keyword Search
- Approximate String Search
- Section Based Parsing and Merging
- Prefix Search
- Associative Keyword Search
- Map-Reduce Parsing

- Efficient
- Effective

Query Times

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Full Craigslist</th>
<th>Full LinkedIn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 200</td>
<td>0.0286</td>
<td>0.0669</td>
</tr>
<tr>
<td>Top [D]</td>
<td>0.0353</td>
<td>0.0889</td>
</tr>
</tbody>
</table>

Design of Entire System

This figure describes the system as a whole which is largely about the construction of the system

A Typical Search

LinkedIn Search

File Triplet

Search 1.7 trillion records in fractions of a second
Parse 1.7 trillion records in less than an hour

Design of Searcher

This figure describes how a search works