Clinical Text Search Engine

Min Jiang
The University of Texas at Houston
01/06/2016
Introduction

• Purpose
  – Help users rapidly locate useful information from clinical text for multiple purposes.

• Plan
  – Prototype (using elastic search)
    • Simple web based application to implement basic functionalities
  – Further optimization and improvement
Elastic search

• Open sourced full-text search engine tool
  – Built on top of Lucene
  – RESTfull service
  – Distributed, scalable and highly available
  – Information stored as JSON documents
  – Sophisticated analytics capabilities
Indexes created by elastic search

<table>
<thead>
<tr>
<th>health</th>
<th>status</th>
<th>index</th>
<th>pri</th>
<th>rep</th>
<th>docs.count</th>
<th>docs.deleted</th>
<th>store.size</th>
<th>pri.store.size</th>
</tr>
</thead>
<tbody>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.10.05</td>
<td>1</td>
<td>1</td>
<td>11607</td>
<td>0</td>
<td>11.2mb</td>
<td>11.2mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>bank</td>
<td>5</td>
<td>1</td>
<td>1000</td>
<td>0</td>
<td>417.4kb</td>
<td>417.4kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.09.29</td>
<td>1</td>
<td>1</td>
<td>20909</td>
<td>0</td>
<td>20.5mb</td>
<td>20.5mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>twitter</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>8.9kb</td>
<td>8.9kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>articles</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>648b</td>
<td>648b</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>customer</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3.2kb</td>
<td>3.2kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.10.19</td>
<td>1</td>
<td>1</td>
<td>6974</td>
<td>0</td>
<td>5.7mb</td>
<td>5.7mb</td>
</tr>
<tr>
<td>red</td>
<td>open</td>
<td>pdb_dev</td>
<td>5</td>
<td>1</td>
<td>909</td>
<td>0</td>
<td>1.9mb</td>
<td>1.9mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.08.10</td>
<td>1</td>
<td>1</td>
<td>12803</td>
<td>0</td>
<td>27.1kb</td>
<td>27.1kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>genenetwork</td>
<td>5</td>
<td>1</td>
<td>66</td>
<td>0</td>
<td>8.9kb</td>
<td>8.9kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.09.16</td>
<td>1</td>
<td>1</td>
<td>4399</td>
<td>0</td>
<td>4.2mb</td>
<td>4.2mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.10.15</td>
<td>1</td>
<td>1</td>
<td>396</td>
<td>0</td>
<td>1.6mb</td>
<td>1.6mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>library</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>7.6kb</td>
<td>7.6kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>sra</td>
<td>5</td>
<td>1</td>
<td>43086</td>
<td>0</td>
<td>7.2mb</td>
<td>7.2mb</td>
</tr>
<tr>
<td>red</td>
<td>open</td>
<td>test</td>
<td>1</td>
<td>1</td>
<td>25142</td>
<td>0</td>
<td>21.9mb</td>
<td>21.9mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.10.12</td>
<td>1</td>
<td>1</td>
<td>8380</td>
<td>0</td>
<td>8.3mb</td>
<td>8.3mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.09.21</td>
<td>1</td>
<td>1</td>
<td>1505</td>
<td>0</td>
<td>1.8mb</td>
<td>1.8mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>note</td>
<td>5</td>
<td>1</td>
<td>4999</td>
<td>0</td>
<td>26.4mb</td>
<td>26.4mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>repository</td>
<td>5</td>
<td>1</td>
<td>537</td>
<td>0</td>
<td>505.4kb</td>
<td>505.4kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.10.27</td>
<td>1</td>
<td>1</td>
<td>877</td>
<td>0</td>
<td>1.3mb</td>
<td>1.3mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>dbgap</td>
<td>5</td>
<td>1</td>
<td>66</td>
<td>0</td>
<td>21.1kb</td>
<td>21.1kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>gwas3</td>
<td>5</td>
<td>1</td>
<td>13198</td>
<td>0</td>
<td>27.6mb</td>
<td>27.6mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-kibana</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3.2kb</td>
<td>3.2kb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.10.09</td>
<td>1</td>
<td>1</td>
<td>18077</td>
<td>0</td>
<td>17.4mb</td>
<td>17.4mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.08.11</td>
<td>1</td>
<td>1</td>
<td>14981</td>
<td>0</td>
<td>19mb</td>
<td>19mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>geo</td>
<td>5</td>
<td>1</td>
<td>3666</td>
<td>0</td>
<td>5.5mb</td>
<td>5.5mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>mtsamples</td>
<td>5</td>
<td>1</td>
<td>4999</td>
<td>0</td>
<td>26.3mb</td>
<td>26.3mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>gwas</td>
<td>5</td>
<td>1</td>
<td>13196</td>
<td>0</td>
<td>2.7mb</td>
<td>2.7mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.08.14</td>
<td>1</td>
<td>1</td>
<td>629</td>
<td>0</td>
<td>1.2mb</td>
<td>1.2mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>.marvel-2015.08.12</td>
<td>1</td>
<td>1</td>
<td>7668</td>
<td>0</td>
<td>10.1mb</td>
<td>10.1mb</td>
</tr>
<tr>
<td>yellow</td>
<td>open</td>
<td>pdb</td>
<td>5</td>
<td>1</td>
<td>112125</td>
<td>0</td>
<td>283.5mb</td>
<td>283.5mb</td>
</tr>
</tbody>
</table>
Sample Type / Medical Specialty: Consult - History and Phy.
Sample Name: Consult - Prostate cancer screening.

options including radical prostatectomy with or without nerve sparing were discussed with him with the risks of bleed or the HDR radiation treatments were discussed with the risks of bladder, bowel, and rectal injury and possible impot impotence. Potency may not recover after the hormone therapy has been completed. Cryosurgery was discussed with the self-intermittent catheterization if he cannot void adequately. Prostate surgery to relieve obstruction and retention understands the options that are available. I spoke with the patient for over 60 minutes concerning these options.

therapy, radical, rectal, rectal fistula, rectal injury, prostate cancer consult, cancer, radiation, prostate,

Sample Type / Medical Specialty: Hematology - Oncology
Sample Name: Colon Cancer Screening.

HISTORY OF PRESENT ILLNESS: Mr. A is a 53-year-old gentleman who was referred for colon cancer screening at all. He denies any hypertension, diabetes, or any other problems. He does not take any medications.

no history of any colon cancer in the family.

REVIEW OF SYSTEMS: Denies any significant diarrhea. Sometimes he gets difficulty in swallowing. Has occasional heartburn.


benefits and risks, in particular the risk of perforation, hemorrhage, and infection. The patient agreed for it. We colonoscopy,
Prototype

• Data collection
  – MT sample notes (5,000)
    • 40 types and specialties

• Build index
  – Elastic search index against full text of clinical note

• Web application development
  – PHP, Javascript
Future Plan

• Further improvements
  – Query expansion on search input
  – Refined ranking algorithm to improve user experience
  – ...

DEMO