

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

health	status	index	pri	rep	docs.count	docs.deleted	store.size	pri.store.size
yellow	open	.marvel-2015.10.05	1	1	11607	0	11.2mb	11.2mb
yellow	open	bank	5	1	1000	0	417.4kb	417.4kb
yellow	open	.marvel-2015.09.29	1	1	20909	0	20.5mb	20.5mb
yellow	open	twitter	5	1	3	0	8.9kb	8.9kb
yellow	open	articles	5	1	0	0	648b	648b
yellow	open	customer	5	1	1	0	3.2kb	3.2kb
yellow	open	.marvel-2015.10.19	1	1	6974	0	5.7mb	5.7mb
red	open	pdb_dev	5	1				
yellow	open	.marvel-2015.08.10	1	1	909	0	1.9mb	1.9mb
yellow	open	genenetwork	5	1	66	0	27.1kb	27.1kb
yellow	open	.marvel-2015.09.16	1	1	12803	0	13.6mb	13.6mb
yellow	open	kimchy	5	1	3	0	8.9kb	8.9kb
yellow	open	.marvel-2015.10.15	1	1	4399	0	4.2mb	4.2mb
yellow	open	.marvel-2015.10.13	1	1	396	0	1.6mb	1.6mb
yellow	open	library	5	1	2	0	7.6kb	7.6kb
yellow	open	sra	5	1	43086	0	7.2mb	7.2mb
red	open	test	1	1				
yellow	open	.marvel-2015.10.14	1	1	25142	0	21.9mb	21.9mb
yellow	open	.marvel-2015.10.12	1	1	8380	0	8.3mb	8.3mb
yellow	open	.marvel-2015.09.21	1	1	1505	0	1.8mb	1.8mb
yellow	open	note	5	1	4999	0	26.4mb	26.4mb
yellow	open	repository	5	1	537	0	505.4kb	505.4kb
yellow	open	.marvel-2015.10.27	1	1	877	0	1.3mb	1.3mb
yellow	open	dbgap	5	1	66	0	21.1kb	21.1kb
yellow	open	gwas3	5	1	13198	0	27.6mb	27.6mb
yellow	open	.marvel-kibana	1	1	1	0	3.2kb	3.2kb
yellow	open	.marvel-2015.10.09	1	1	18077	0	17.4mb	17.4mb
yellow	open	.marvel-2015.08.11	1	1	14981	0	19mb	19mb
yellow	open	geo	5	1	3666	0	5.5mb	5.5mb
yellow	open	mtsamples	5	1	4999	0	26.3mb	26.3mb
yellow	open	gwas	5	1	13196	0	2.7mb	2.7mb
yellow	open	.marvel-2015.08.14	1	1	629	0	1.2mb	1.2mb
yellow	open	.marvel-2015.08.12	1	1	7668	0	10.1mb	10.1mb
yellow	open	pdb	5	1	112125	0	283.5mb	283.5mb

Elastic Search Result

```
{
  "took" : 13,
  "timed_out" : false,
  "_shards" : {
    "total" : 5,
    "successful" : 5,
    "failed" : 0
  },
  "hits" : {
    "total" : 480,
    "max_score" : 0.48834965,
    "hits" : [ {
      "_index" : "note",
      "_type" : "search",
      "_id" : "774",
      "_score" : 0.48834965,
      "_source":{"text": "Sample Type / Medical Specialty: Consult - History and Phy.\nSample Name: Consult - Prosta
options including radical prostatectomy with or without nerve sparing were discussed with him with the risks of blee
or the HDR radiation treatments were discussed with the risks of bladder, bowel, and rectal injury and possible impot
impotency. 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 retentic
understands the options that are available. I spoke with the patient for over 60 minutes concerning these options.\n
therapy, radical, rectal, rectal fistula, rectal injury, prostate cancer consult, cancer, radiation, prostate,\n", "t
    }, {
      "_index" : "note",
      "_type" : "search",
      "_id" : "1899",
      "_score" : 0.47626495,
      "_source":{"text": "Sample Type / Medical Specialty: Hematology - Oncology\nSample Name: Colon Cancer Screeni
colon cancer screening.\nHISTORY OF PRESENT ILLNESS: Mr. A is a 53-year-old gentleman who was referred for colon can
problems at all. He denies any hypertension, diabetes, or any other problems. He does not take any medications.\nP
no history of any colon cancer in the family.\nREVIEW OF SYSTEMS: Denies any significant diarrhea. Sometimes he gets
difficulty in swallowing. Has occasional heartburn.\nPHYSICAL EXAMINATION: The patient is alert and oriented x3. V:
heard. Rhythm is regular. No murmur. Lungs: Clear to percussion and auscultation. Abdomen: Soft and nontender. 1
benefits and risks, in particular the risk of perforation, hemorrhage, and infection. The patient agreed for it. We
colonoscopy,\n", "title": "Hematology_Oncology/sample_2793.txt"}
```

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