MediCrawl - A Web Search Engine For Medical Diagnosis

Devharsh Trivedi,[1] Vaishnavi Gopalakrishnan[2] and Wendy Hui Wang[3]

Abstract— **Popular generic web search engines like Google and Bing fail to be as effective as niche web search engine for searches related to a specific domain. There are solutions available like MayoClinic and FindZebra for searching medical diseases based on symptoms but they have their limitations. We present in this paper MediCrawl - A web search engine made with open source tools for searching diseases. In this project, we target both the common and the rare diseases and try to overcome the limitations imposed by like MayoClinic and FindZebra.**

I. INTRODUCTION

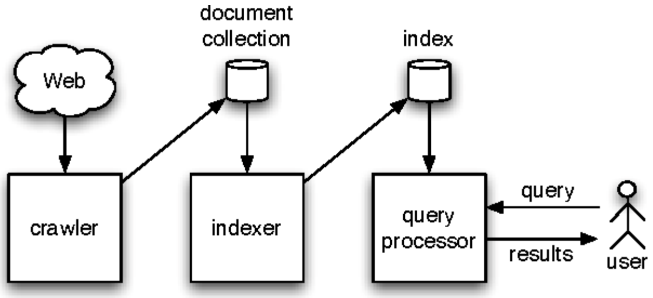


Fig. 1: Components of a web search engine

A. *Web Search Engine*

Internet is our everyday source for all kinds of information, like news, broadcast, music, pictures, or even movies. We are immersed in the flood of big data every day. However, the information we need is only a small portion of it and for different individuals the importance of information is hardly the same. A search engine will make it convenient for us to absorb information efficiently.

B. *Crawling and Indexing*

Search Engines have automated robots called crawlers that use web links to scour the Internet, find web pages, and decipher page data that are indexed to be included in search engine results.

C. *Ranking*

Each search engine has different ranking factors, but they all have a few factors in common: keywords, content, and links. Keywords and content are arguably two of the most important factors that search engines look for when ranking pages. Because of this, it is imperative that you know what keywords are in highest demand within your market and incorporate those keywords into the content on your website. All your website's content will naturally create a collection of links; search engines use link analysis algorithms that look at the sources, number, and anchor texts of links to help determine their relevance in search queries. TF-IDF based retrieval is one such method for ranking.

II. SCOPE OF WORK

A. *Problem Definition Related Work*

We have niche web search engine for medical diagnosis like MayoClinic and FindZebra which are quite popular. MayoClinic [1] is useful when you want to get more details about a particular disease and it also provides a symptom checker tool when you don't know about the disease but want to run a diagnose based on the symptoms. Problem with this kind of approach is that it does not give you flexibility of free text search and you must rely upon the given parameters and choose the relevant ones from the list to diagnose. There is another solution called FindZebra [2] which aims at diagnosing of rare diseases. It has a free search capability where you can type in the symptoms and it will retrieve the ranked searches based on the user query. FindZebra is targeted for rare diseases and does not work well for common diseases. To tackle these problems, we present in this paper our solution "MediCrawl" - A web search engine for any type of diseases. In this project we target both the common and the rare diseases in children and adults and provide

[1] Devharsh: He is a Ph.D. student in the Department of Computer Science, Stevens Institute of Technology, NJ Email: dtrived5@stevens.edu

[2] Vaishnavi: She is a Masters student in the Department of Computer Science, Stevens Institute of Technology, NJ Email: vgopalak@stevens.edu

[3] Wendy: She is an Associate Professor in the Department of Computer Science, Stevens Institute of Technology, NJ Email: hwang4@stevens.edu

free text search capability. This solution can be used by everyone regardless of their medical expertise.

B. Vision

We have used open-source libraries to build our web crawler to download the data from the internet, and we are using No SQL database to store the data retrieved by the crawler, and we have built a simple web interface as front-end for users to search their query using HTML5, CSS3, JavaScript, jQuery & Ajax. Java Run-time Environment is used.

C. Accessibility and Indexing

For the website to be visible in search results, search engines must be able to find it. They do this by crawling the web and looking for relevant and index-able content, such as link structures and HTML features like alt tags for images.

D. Ranking Factors

The different factors that we considered are 1) contents, to make sure that your content is relevant and utilizes keywords without being duplicated on multiple pages, 2) keywords, and 3) HTML, to make sure that your source code is relevant to crawlers.

E. Data Management

The data management plan was done to get the data for our website. We implemented a search engine by building our own indexed database to collect web pages via Web crawling.

F. Query Implementation

Our requirements impose that query results get displayed within 400ms. And the relevance (confidence of retrieved results) should be more than 75%.

G. Major Contributions

We have used free and open source tools and technologies to

- Originate a ground truth of symptoms and diseases based on trusted data sources which can be used to evaluate the performance of a web search engine.
- Discover a list of trustworthy web-pages that serves as starting point of our Crawler.
- Generate our own database of inverted indexed using our Crawler and store it in a No-SQL like data-store.
- Implement a ranking and re-ranking strategy to retrieve accurate results for a given query.
- Design a user interface that is intuitive and easy to use and which helps a user to get the best information without having to leave the web page.

H. Role of members

Devharsh (lead)

1) Project planning and design - 1.5 hours
2) Documentation for project proposal - 1 hour
3) Identifying list of seed websites to crawl and establishing ground truth to evaluate query retrieval performance - 4 hours
4) Implement web spider to crawl through several levels deep from the starting pages - 4 hours
5) Implement web server database to store inverted index - 4 hours
6) Crawling and Indexing to build own database - 16 hours
7) Documentation for midterm report - 2.5 hours
8) Final project presentation - 2 hours
9) Testing and Evaluation: MRR, P@10, P@20 - 4.5 hours
10) Final project report - 6.5 hours

Vaishnavi

1) Project conception and idealization - 1 hour
2) Documentation for project proposal - 2 hours
3) Implement free text search user interface - 3 hours
4) Implement HTTP request API and displaying results - 3.5 hours
5) Implement Ranking (BM25) and Re-ranking (LTR Linear Model) - 9 hours
6) Documentation for midterm report - 2 hours
7) Implement Wikipedia API - 4 hours
8) Implement Google API - 4 hours
9) Final project presentation - 2 hours
10) Final project report - 3.5 hours

TABLE I: Project Schedule

Module	Planned	Progress	Member
Planning	March 15	Completed	Devharsh
Design	March 28	Completed	Vaishnavi
Crawling	April 5	Completed	Devharsh
Indexing	April 10	Completed	Vaishnavi
Ranking	April 20	Completed	Devharsh
Searching	May 4	Completed	Vaishnavi
Testing	May 8	Completed	Both
Improvements	May 9	Completed	Devharsh
Evaluation	May 10	Completed	Devharsh
Documentation	May 12	Completed	Both

TABLE II: Tools & Technologies

Software	Functionality	Module Type	Link
Java	Runtime Environment	Development platform	https://www.java.com/en/
Git	Source code management	Version control	https://git-scm.com/
Git-LFS	Source code management	Version Control	https://git-lfs.github.com/
Nutch	Crawling/Indexing	Web Crawler	http://nutch.apache.org/downloads.html
Solr	Indexing/Ranking	Database	https://lucene.apache.org/solr/downloads.html
HTML/CSS/jQuery	Searching	User Interface	https://html-css-js.com/

III. Implementation

We have implemented a web crawler to search and retrieve the documents containing the user query for diseases and indexed the retrieved documents in Solr, a No SQL database, which can hold a large set of results. A clean and easy to use user interface was designed to facilitate three types of searches (i) search from our database (ii) Wikipedia search and (iii) customised Google search. The results retrieved from our database are ranked using BM25 and LinearModel.

A. *Crawling & Indexing*

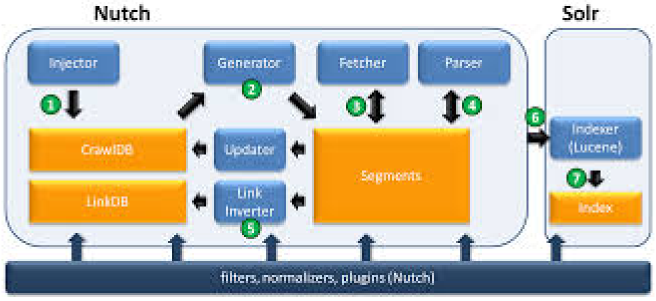


Fig. 2: Crawling & Indexing with Nutch & Solr

We used Solr along with Nutch to index the crawled web pages in the database. Solr is the third most popular database used, after ElasticSearch and Splunk.[3]

a) Apache Nutch: [4]

- Nutch is an open source search engine implemented in Java.
- Nutch implements "Map Reduce" distributed processing model.
- Nutch installations typically operate at one of three scales: local file system, intranet, or whole web.
- Nutch is built on top of Lucene, which is an API for text indexing and searching.
- Configuration:
  - nutch/conf/nutch-site.xml
  - nutch/conf/schema.xml
  - nutch/urls/seed.txt
- Crawl and Index simultaneously: nutch/bin/crawl -i -Dsolr.server.url=http://localhost:8983/solr/nutch -s nutch/urls/ Crawl 3

b) Apache Lucene: [5]

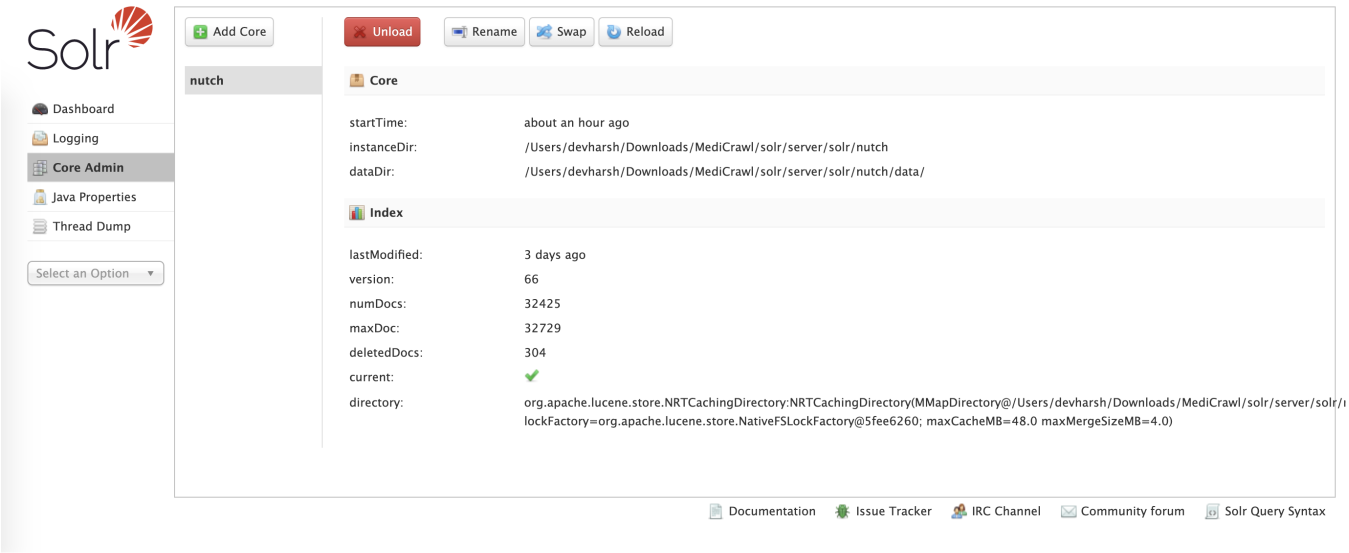
- NOT a crawler
- NOT an application
- NOT a library for doing Google page rank
- An open source Java-based Information Retrieval library enabling text based search

c) Nutch v/s Lucene: [4]

- Use Lucene, if a web crawler is not needed.
- A common scenario is that you have a web front end to a database that you want to make search able. The best way to do this is to index the data directly from the database using the Lucene API, and then write code to do searches against the index, again using Lucene.
- Nutch is a better fit for sites where you don't have direct access to the underlying data, or it comes from disparate sources.

d) Nutch Architecture: [4]

- Nutch is divided into two pieces: the crawler and the searcher.
- The crawler fetches pages and turns them into an inverted index.
- This inverted index is used by the searcher to resolve user's queries.
- Searcher and crawler components can be scaled independently of each other.
- The crawler system is driven by:
  - Nutch crawl tool.
  - Several types of data structures, including the web database, a set of segments, and the index.
- The web database, or WebDB stores two types of entities: pages and links.



startTime: about an hour ago
instanceDir: /Users/devharsh/Downloads/MediCrawl/solr/server/solr/nutch
dataDir: /Users/devharsh/Downloads/MediCrawl/solr/server/solr/nutch/data/

lastModified: 3 days ago
version: 66
numDocs: 32425
maxDoc: 32729
deletedDocs: 304
current: ✓
directory: org.apache.lucene.store.NRTCachingDirectory:NRTCachingDirectory(MMapDirectory@/Users/devharsh/Downloads/MediCrawl/solr/server/solr/i
lockFactory=org.apache.lucene.store.NativeFSLockFactory@5fee6260; maxCacheMB=48.0 maxMergeSizeMB=4.0)

Fig. 3: Web pages crawled using Nutch and Indexed in Solr

- A page represents a page on the Web and is indexed by its URL and the MD5 hash of its contents.
- A link represents a link from one web page (the source) to another (the target). A segment is a collection of pages fetched and indexed by the crawler in a single run.
- The fetch-list for a segment is a list of URLs for the crawler to fetch and is generated from the WebDB.
- The fetcher output is the data retrieved from the pages in the fetch-list. The fetcher output for the segment is indexed and the index is stored in the segment.
- The index is the inverted index of all the pages the system has retrieved and is created by merging all of the individual segment indexes.
- Nutch uses Lucene for its indexing.

e) Steps performed by a Crawler: [4]

1) Create a new WebDB.
2) Inject root URLs into the WebDB.
3) Generate a fetch-list from the WebDB in a new segment.
4) Fetch content from URLs in the fetch-list.
5) Update the WebDB with links from fetched pages.
6) Repeat steps 3-5 until the required depth is reached.
7) Update segments with scores and links from the WebDB.
8) Index the fetched pages.
9) Eliminate duplicate content (and duplicate URLs) from the indexes.
10) Merge the indexes into a single index for searching.

f) Apache Solr: [5]

- An open source enterprise search server
- Based on the Lucene Java search library
- A web-based application that processes HTTP request and returns HTTP responses
- Completed with XML/HTTP APIs, caching, replication, and web administration interface.

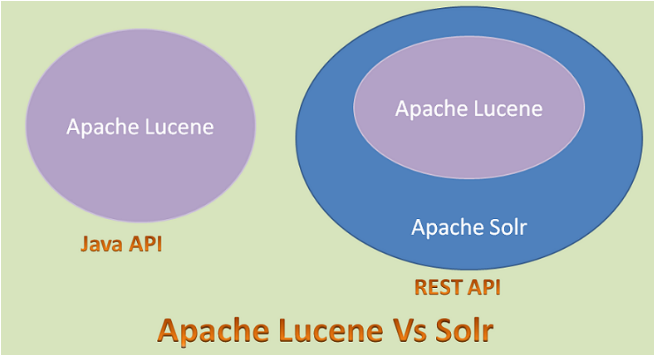


Fig. 4: Apache Solr and Lucene Libraries

g) Solr as NoSQL: [5]

- NoSQL : No defined structure for the database
- Characteristics:
  - Non-traditional data stores
  - Not designed for SQL type query

  - Document oriented, data format agnostic (JSON, XML, CSV, binary)
- Versioning and optimistic locking with Real Time GET, allows read/write/update with/without conflict.
- Atomic updates can add/remove/change and increment a field in existing index with/without re-indexing
- Create a table(core) in Solr before indexing from Nutch
- Configuration: solr/server/solr/configsets/nutch/conf/ solrconfig.xml
- Run: solr/bin/solr start -Dsolr.ltr.enabled=true

B. Ranking

In information retrieval systems, Learning to Rank is used to re-rank the top N retrieved documents using trained machine learning models. Using such sophisticated models can make more nuanced ranking decisions than standard ranking functions like TF-IDF or BM25. We experimented with popular ranking algorithms like BM25 and Linear Model to determine which works best for our case.

a) BM25: BM25 improves upon TF*IDF. BM25 stands for “Best Match 25”. Released in 1994, it’s the 25th iteration of tweaking the relevance computation. BM25 has its roots in probabilistic information retrieval. Basically, it casts relevance as a probability problem. A relevance score, according to probabilistic information retrieval, ought to reflect the probability a user will consider the result relevant.

Sample response showing BM25 score:

```
{
 "responseHeader":{
  "status":0,
  "QTime":31,
  "params":{
   "q":"breast cancer",
   "fl":"title,url,score,
   [features]"}},
 "response":{"numFound":6853,
 "start":0,"maxScore":4.4154315,
 "docs":[
   {
    "title":"Breast Cancer",
    "url":"https://www.pinterest
    .com/amp/cdcgov/breast-cancer/",
    "score":4.4154315,
    "[features]":
    "originalScore=4.4154315,
    titleLength=2.0,
    contentLength=7192.0"},
    .
    .
```

b) LinearModel: We are using LinearModel which is a scoring model that computes scores using a dot product.[6] We have used following model:

```
{
 "responseHeader":{
  "status":0,
  "QTime":139},
 "models":[{
   "name":"myModel",
   "class":"org.apache.solr.ltr
   .model.LinearModel",
   "store":"_DEFAULT_",
   "features":[{
     "name":"contentLength",
     "norm":{"class":"org.apache
     .solr.ltr.norm
     .IdentityNormalizer"}},
    {
     "name":"titleLength",
     "norm":{"class":"org.apache
     .solr.ltr.norm
     .IdentityNormalizer"}},
    {
     "name":"originalScore",
     "norm":{"class":"org.apache
     .solr.ltr.norm
     .IdentityNormalizer"}}],
   "params":{"weights":{
     "contentLength":1.0,
     "titleLength":0.1,
     "originalScore":0.5}}}]}
```

Sample response showing our model based score:

```
  {
 "responseHeader":{
  "status":0,
  "QTime":20,
  "params":{
   "q":"breast cancer",
   "fl":"title,url,score,
   [features]",
   "rq":"{!ltr model=myModel
   efi.query=breast cancer}"}},
 "response":{"numFound":6853,"start":0,
 "maxScore":4.4154315,"docs":[
{
"title":"Cancer - Wikipedia",
"url":"https://en.wikipedia.org/
wiki/Cancer",
"score":18458.223,
"[features]":"
originalScore=4.0475817,
titleLength=2.0,
contentLength=18456.0"},
.
.
```

C. Searching in Solr[5]

- The search query is processed by a Request Handler:
  - Request Handler calls a query parser
  - Query parser interprets query’s term & parameters
  - Input to a query parser can include:
    - Search strings – common terms
    - Parameters for fine tuning, e.g. Boolean logic
    - Parameters for controlling the presentation of the query response, e.g. Specifying the order in which results are displayed.
- Searching in Solr can be done by sending HTTP GET and POST requests: http://localhost:8983/solr/select?q=covid-19
- Solr provides a simple method to sort on 1 or more indexed fields. Use the sort parameter: http://localhost:8983/solr/select?q=covid-19 &sort=title asc

D. User Interface

Developed a web-based user interface using HTML5, CSS3, and jQuery. The user interface is composed of a customized google search for diseases. It shows the web pages in Web tab and related images in Image tab. There is a search box implemented to input the user query. The query can be a direct search of any disease name or the symptoms to find all the related diseases. Two tabs viz., MediCrawl and Wikipedia are provided for search. Wikipedia and Google search can be used when the user finds difficult to understand any term displayed in the result set by the MediCrawl search. They do not have to jump to another website or new tab in the browser to search for unknown terms. The pagination is done based on the number of documents available in the database, each page displaying a maximum of 25 results.

E. Procedure

Perform the following steps to setup the web search engine from scratch. Make sure to install prerequisites software.

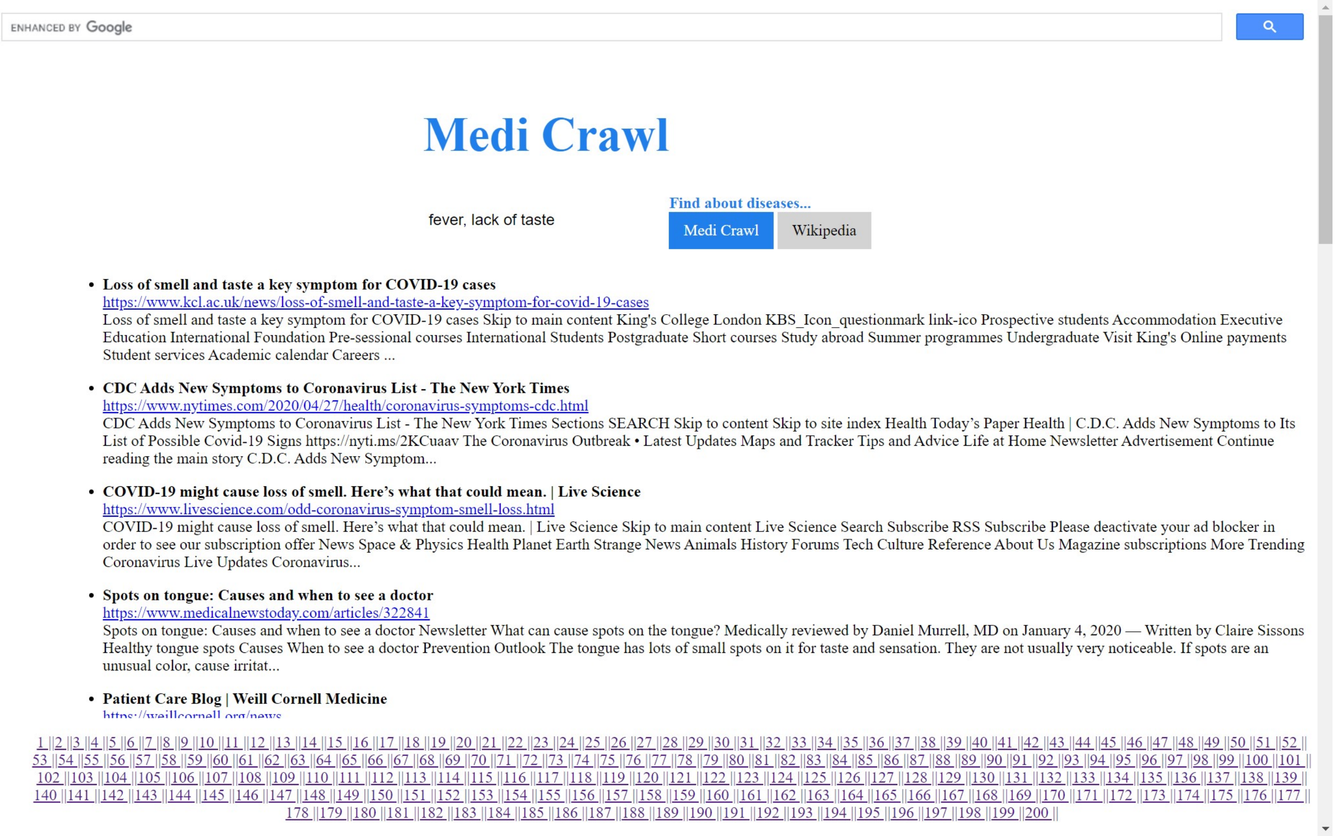
1) Download and extract Nutch and Solr in nutch and solr directories
2) Modify nutch/conf/nutch-site.xml: set http.agent.name and indexr-solr property
3) Modify nutch/conf/schema.xml
4) Create a URL seed list: nutch/urls/seed.txt
5) cp -r solr/server/solr/configsets/_default solr/server/solr/configsets/nutch
6) cp nutch/conf/schema.xml solr/server/solr/configsets/nutch/conf
7) rm solr/server/solr/configsets/nutch/conf/managed-schema
8) Modify solr/server/solr/configsets/nutch/conf/solrconfig .xml to enable LTR and change _text_ to text
9) start solr with LTR: solr/bin/solr start -Dsolr.ltr.enabled=true
10) set JAVA_HOME environment variable
11) Crawl and Index simultaneously: nutch/bin/crawl -i -D solr.server.url=http://localhost:8983/solr/nutch -s nutch/urls/ Crawl 3
12) Test if it succeeded: curl ’http://localhost:8983/solr/nutch/query?q=breast+cancer &fl=url,title’
13) Create /path/features.json and /path/model.json
14) Upload features to Solr: curl -XPUT ’http://localhost:8983/solr/nutch/schema/feature-store’ –data-binary ”@./path/features.json” -H ’Content-type:application/json’
15) View uploaded features: curl ’http://localhost:8983/solr/nutch/schema/feature-store/ _DEFAULT_’
16) Test feature extraction: curl ’http://localhost:8983/solr/nutch/query?q=test&fl=id, score,[features]’
17) Upload model to Solr: curl -XPUT ’http://localhost:8983/solr/nutch/schema/model-store’ –data-binary ”@./path/model.json” -H ’Content-type:application/json’
18) View uploaded model: curl ’http://localhost:8983/solr/nutch/schema/model-store’
19) Test model: curl -g ’http://localhost:8983/solr/nutch/query?q=breast+cancer &rq={!ltr%20model=myModel%20efi.query=breast+ cancer}&fl=url,title,[features]’

F. Sample Responses

Following are the results as JSON responses you will get based on the query performed.

a) A Basic Query for ”Breast Cancer”:

```
{
  "responseHeader":{
```



- **Loss of smell and taste a key symptom for COVID-19 cases**
  https://www.kcl.ac.uk/news/loss-of-smell-and-taste-a-key-symptom-for-covid-19-cases
  Loss of smell and taste a key symptom for COVID-19 cases Skip to main content King's College London KBS_Icon_questionmark link-ico Prospective students Accommodation Executive Education International Foundation Pre-sessional courses International Students Postgraduate Short courses Study abroad Summer programmes Undergraduate Visit King's Online payments Student services Academic calendar Careers ...

- **CDC Adds New Symptoms to Coronavirus List - The New York Times**
  https://www.nytimes.com/2020/04/27/health/coronavirus-symptoms-cdc.html
  CDC Adds New Symptoms to Coronavirus List - The New York Times Sections SEARCH Skip to content Skip to site index Health Today's Paper Health | C.D.C. Adds New Symptoms to Its List of Possible Covid-19 Signs https://nyti.ms/2KCuaav The Coronavirus Outbreak • Latest Updates Maps and Tracker Tips and Advice Life at Home Newsletter Advertisement Continue reading the main story C.D.C. Adds New Symptom...

- **COVID-19 might cause loss of smell. Here’s what that could mean. | Live Science**
  https://www.livescience.com/odd-coronavirus-symptom-smell-loss.html
  COVID-19 might cause loss of smell. Here’s what that could mean. | Live Science Skip to main content Live Science Search Subscribe RSS Subscribe Please deactivate your ad blocker in order to see our subscription offer News Space & Physics Health Planet Earth Strange News Animals History Forums Tech Culture Reference About Us Magazine subscriptions More Trending Coronavirus Live Updates Coronavirus...

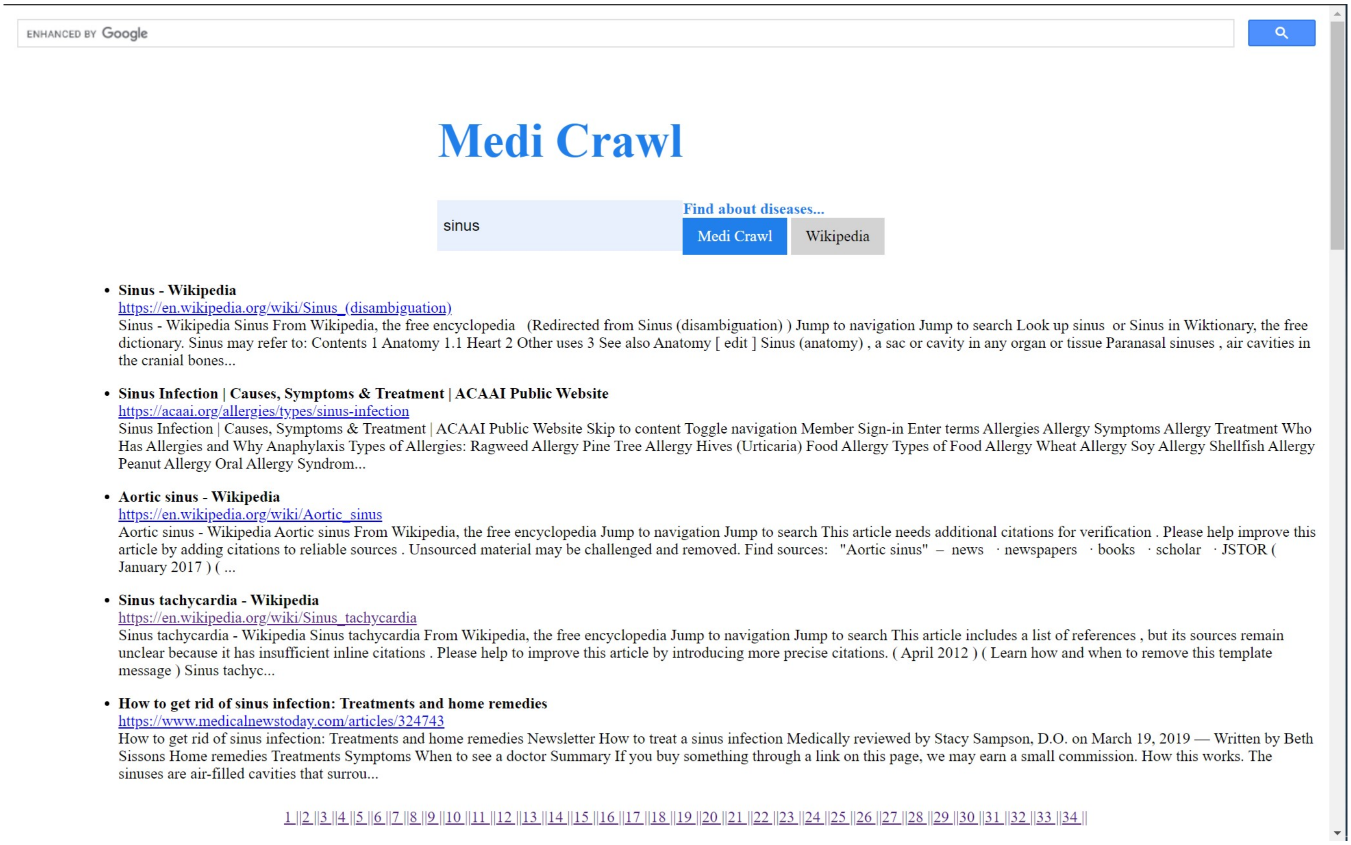
- **Spots on tongue: Causes and when to see a doctor**
  https://www.medicalnewstoday.com/articles/322841
  Spots on tongue: Causes and when to see a doctor Newsletter What can cause spots on the tongue? Medically reviewed by Daniel Murrell, MD on January 4, 2020 — Written by Claire Sissons Healthy tongue spots Causes When to see a doctor Prevention Outlook The tongue has lots of small spots on it for taste and sensation. They are not usually very noticeable. If spots are an unusual color, cause irritat...

- **Patient Care Blog | Weill Cornell Medicine**
  https://weillcornell.org/news

1 ||2 ||3 ||4 ||5 ||6 ||7 ||8 ||9 ||10 ||11 ||12 ||13 ||14 ||15 ||16 ||17 ||18 ||19 ||20 ||21 ||22 ||23 ||24 ||25 ||26 ||27 ||28 ||29 ||30 ||31 ||32 ||33 ||34 ||35 ||36 ||37 ||38 ||39 ||40 ||41 ||42 ||43 ||44 ||45 ||46 ||47 ||48 ||49 ||50 ||51 ||52 ||
53 ||54 ||55 ||56 ||57 ||58 ||59 ||60 ||61 ||62 ||63 ||64 ||65 ||66 ||67 ||68 ||69 ||70 ||71 ||72 ||73 ||74 ||75 ||76 ||77 ||78 ||79 ||80 ||81 ||82 ||83 ||84 ||85 ||86 ||87 ||88 ||89 ||90 ||91 ||92 ||93 ||94 ||95 ||96 ||97 ||98 ||99 ||100 ||101 ||
102 ||103 ||104 ||105 ||106 ||107 ||108 ||109 ||110 ||111 ||112 ||113 ||114 ||115 ||116 ||117 ||118 ||119 ||120 ||121 ||122 ||123 ||124 ||125 ||126 ||127 ||128 ||129 ||130 ||131 ||132 ||133 ||134 ||135 ||136 ||137 ||138 ||139 ||
140 ||141 ||142 ||143 ||144 ||145 ||146 ||147 ||148 ||149 ||150 ||151 ||152 ||153 ||154 ||155 ||156 ||157 ||158 ||159 ||160 ||161 ||162 ||163 ||164 ||165 ||166 ||167 ||168 ||169 ||170 ||171 ||172 ||173 ||174 ||175 ||176 ||177 ||
178 ||179 ||180 ||181 ||182 ||183 ||184 ||185 ||186 ||187 ||188 ||189 ||190 ||191 ||192 ||193 ||194 ||195 ||196 ||197 ||198 ||199 ||200 ||

Fig. 5: Web User Interface - Symptom search

```
"status":0,
"QTime":0,
"params":{
  "q":"breast cancer",
  "fl":"url,title"}},
"response":{"numFound":1122,"start":0,
"docs":[
  {
    "title":"Breast cancer - Symptoms and
    causes - Mayo Clinic",
    "url":"https://www.mayoclinic.org/
    diseases-conditions/breast-cancer/
    symptoms-causes/syc-20352470"},
  {
    "title":"Breast cancer - Diagnosis and
    treatment - Mayo Clinic",
    "url":"https://www.mayoclinic.org/
    diseases-conditions/breast-cancer/
    diagnosis-treatment/drc-20352475"},
  {
    "title":"Breast cancer - Care at Mayo
    Clinic - Mayo Clinic",
    "url":"https://www.mayoclinic.org/
    diseases-conditions/breast-cancer/
    care-at-mayo-clinic/mac-20352479"},
  {
    "title":"Breast cancer: Symptoms,
    causes, and treatment",
    "url":"https://www.medicalnewstoday.com/
    articles/37136"},
  {
    "title":"Breast Clinic - Overview -
    Mayo Clinic",
    "url":"https://www.mayoclinic.org/
    departments-centers/breast-clinic/
    sections/overview/ovc-20459469"},
  {
    "title":"Bring Your Brave.",
    "url":"https://bringyourbrave.tumblr.
    com/"},
  {
    "title":"Breast Cancer News from
    Medical News Today",
    "url":"https://www.medicalnewstoday.com/
    categories/breast-cancer"},
  {
    "title":"Breast Cancer | CDC",
    "url":"https://www.cdc.gov/cancer/
    breast/"},
  {
    "title":"Breast cancer - Doctors and
    departments - Mayo Clinic",
```



- **Sinus - Wikipedia**
  https://en.wikipedia.org/wiki/Sinus_(disambiguation)
  Sinus - Wikipedia Sinus From Wikipedia, the free encyclopedia (Redirected from Sinus (disambiguation)) Jump to navigation Jump to search Look up sinus or Sinus in Wiktionary, the free dictionary. Sinus may refer to: Contents 1 Anatomy 1.1 Heart 2 Other uses 3 See also Anatomy [edit] Sinus (anatomy) , a sac or cavity in any organ or tissue Paranasal sinuses , air cavities in the cranial bones...
- **Sinus Infection | Causes, Symptoms & Treatment | ACAAI Public Website**
  https://acaai.org/allergies/types/sinus-infection
  Sinus Infection | Causes, Symptoms & Treatment | ACAAI Public Website Skip to content Toggle navigation Member Sign-in Enter terms Allergies Allergy Symptoms Allergy Treatment Who Has Allergies and Why Anaphylaxis Types of Allergies: Ragweed Allergy Pine Tree Allergy Hives (Urticaria) Food Allergy Types of Food Allergy Wheat Allergy Soy Allergy Shellfish Allergy Peanut Allergy Oral Allergy Syndrom...
- **Aortic sinus - Wikipedia**
  https://en.wikipedia.org/wiki/Aortic_sinus
  Aortic sinus - Wikipedia Aortic sinus From Wikipedia, the free encyclopedia Jump to navigation Jump to search This article needs additional citations for verification . Please help improve this article by adding citations to reliable sources . Unsourced material may be challenged and removed. Find sources: "Aortic sinus" – news · newspapers · books · scholar · JSTOR (January 2017) (...
- **Sinus tachycardia - Wikipedia**
  https://en.wikipedia.org/wiki/Sinus_tachycardia
  Sinus tachycardia - Wikipedia Sinus tachycardia From Wikipedia, the free encyclopedia Jump to navigation Jump to search This article includes a list of references , but its sources remain unclear because it has insufficient inline citations . Please help to improve this article by introducing more precise citations. (April 2012) (Learn how and when to remove this template message) Sinus tachyc...
- **How to get rid of sinus infection: Treatments and home remedies**
  https://www.medicalnewstoday.com/articles/324743
  How to get rid of sinus infection: Treatments and home remedies Newsletter How to treat a sinus infection Medically reviewed by Stacy Sampson, D.O. on March 19, 2019 — Written by Beth Sissons Home remedies Treatments Symptoms When to see a doctor Summary If you buy something through a link on this page, we may earn a small commission. How this works. The sinuses are air-filled cavities that surrou...

1 ||2 ||3 ||4 ||5 ||6 ||7 ||8 ||9 ||10 ||11 ||12 ||13 ||14 ||15 ||16 ||17 ||18 ||19 ||20 ||21 ||22 ||23 ||24 ||25 ||26 ||27 ||28 ||29 ||30 ||31 ||32 ||33 ||34 ||

Fig. 6: Web User Interface - Disease search

```
      "url":"https://www.mayoclinic.org/
      diseases-conditions/breast-cancer/
      doctors-departments/ddc-20352478"},
    {
      "title":"Breast Cancer | Disease of
      the Week | CDC",
      "url":"https://www.cdc.gov/dotw/
      breastcancer/index.html"}]
}}
```

b) Query with Features score for ”Breast Cancer”:

```
  {
"responseHeader":{
  "status":0,
  "QTime":8,
  "params":{
    "q":"breast cancer",
    "fl":"url,title,[features]",
    "rq":"{!ltr model=myModel efi.query=
    breast cancer}"}},
  "response":{"numFound":1122,"start":0,
  "docs":[
    {
      "title":"Science Clips - Volume 12,
      Issue 10, March 23, 2020",
      "url":"https://www.cdc.gov/library/
      sciclips/issues/index.html",
      "[features]":"originalScore=2.2747984,
      titleLength=9.0,contentLength=22552.0"},
    {
      "title":"Science Clips - Volume 12,
      Issue 10, March 23, 2020",
      "url":"https://www.cdc.gov/library/
      sciclips/issues/",
      "[features]":"originalScore=2.2747984,
      titleLength=9.0,contentLength=22552.0"},
    {
      "title":"Breast cancer - Diagnosis
      and treatment - Mayo Clinic",
      "url":"https://www.mayoclinic.org/
      diseases-conditions/breast-cancer/
      diagnosis-treatment/drc-20352475",
      "[features]":
      "originalScore=3.7579455,
      titleLength=7.0,
      contentLength=5144.0"},
    {
      "title":"The Topic Is Cancer | Blogs |
      CDC",
      "url":"https://blogs.cdc.gov/cancer/",
      "[features]":"originalScore=3.5015173,
        titleLength=6.0,
        contentLength=5144.0"},
      {
        "title":"All Issues - Mayo Clinic
        Health Letter",
        "url":"https://healthletter.mayoclinic.
        com/issues",
        "[features]":"originalScore=3.071907,
        titleLength=6.0,contentLength=5144.0"},
      {
        "title":"Menopause Treatment, Signs,
        Symptoms & Age",
        "url":"https://www.medicinenet.com/
        menopause/article.htm",
        "[features]":"originalScore=2.6457264,
        titleLength=5.0,contentLength=5144.0"},
      {
        "title":"Hormone Therapy for Women:
        Side Effects, Cancer Risks",
        "url":"https://www.medicinenet.com/
        hormone_therapy/article.htm",
        "[features]":"originalScore=3.3211832,
        titleLength=8.0,contentLength=4632.0"},
      {
        "title":"Cáncer de mama - Atención en
        Mayo Clinic - Mayo Clinic",
        "url":"https://www.mayoclinic.org/es-es/
        diseases-conditions/breast-cancer/
        care-at-mayo-clinic/mac-20352479",
        "[features]":"originalScore=3.6176624,
        titleLength=9.0,contentLength=4120.0"},
      {
        "title":"Hot Flashes Causes, Symptoms &
        Treatment Medicine for Men & Women",
        "url":"https://www.medicinenet.com/
        hot_flashes/article.htm",
        "[features]":"originalScore=2.8397057,
        titleLength=9.0,contentLength=3864.0"},
      {
        "title":"Breast cancer - Care at Mayo
        Clinic - Mayo Clinic",
        "url":"https://www.mayoclinic.org/
        diseases-conditions/breast-cancer/
        care-at-mayo-clinic/mac-20352479",
        "[features]":"originalScore=3.7574568,
        titleLength=8.0,contentLength=3608.0"}]
  }}
```

c) Pagination - Second page 25 results:

```
{
  "responseHeader":{
    "status":0,
    "QTime":0,
    "params":{
      "q":"lack of taste,fever",
      "fl":"title,url",
      "start":"25",
      "rows":"25"}},
      "response":{"numFound":27832,
      "start":25,"docs":[
      {
        "title":"Somnolence - Wikipedia",
        "url":"https://en.wikipedia.org/wiki/
        Somnolence"},
      {
        "title":"Loss of smell may suggest milder
        COVID-19, study finds",
        "url":"https://www.medicalnewstoday.com/
        articles/loss-of-smell-may-suggest-milder
        -covid-19-study-finds"},
      {
        "title":"Coronavirus (COVID-19) Pandemic:
        What to Do if Your Child Is Sick (for
        Parents) - Nemours KidsHealth",
        "url":"https://kidshealth.org/en/parents
        /coronavirus-child-is-sick.html"},
      {
        "title":"Mom with Multiple Sclerosis
        Overcomes COVID-19 at Home",
        "url":"https://my.clevelandclinic.org/
        patient-stories/376-mom-with-multiple-
        sclerosis-overcomes-covid-19-at-home"},
      {
        "title":"Different Symptoms for
        Coronavirus, Flu, and Allergies",
        "url":"https://www.healthline.com/health
        -news/flu-allergies-coronavirus-different
        -symptoms"},
      {
        "title":"Breastfeeding Isn't a Solo Job
        | How a Partner's Support Is Every",
        "url":"https://www.healthline.com/health
        /parenting/breastfeeding-support"},
      {
        "title":"Sensory processing disorders:
        Definition, symptoms, and more",
        "url":"https://www.medicalnewstoday.com
        /articles/sensory-processing-disorder"},
      {
        "title":"Water Disinfection | Travelers'
        Health | CDC",
        "url":"https://wwwnc.cdc.gov/travel/page
        /water-disinfection"},
      {
        "title":"Coronavirus (COVID-19) (for
        Parents) - Nemours KidsHealth",
        "url":"https://kidshealth.org/en/parents
        /coronavirus.html"},
      {
        "title":"Naegleriasis - Wikipedia",
    "url":"https://en.wikipedia.org/wiki
    /Naegleriasis"},
  {
    "title":"Coronavirus (COVID-19) symptoms:
    How to know and what to do",
    "url":"https://www.medicalnewstoday.com/
    articles/covid-19-symptoms"},
  {
    "title":"Coronavirus symptoms: What are
    they and how do I protect myself?
    - BBC News",
    "url":"https://www.bbc.co.uk/news/health
    -51048366"},
  {
    "title":"Unconsciousness - Wikipedia",
    "url":"https://en.wikipedia.org/wiki/
    Unconsciousness"},
  {
    "title":"Radicalised Youth - Al
    Jazeera English",
    "url":"https://www.aljazeera.com/
    programmes/radicalised-youth/
    radicalised-youth.html"},
  {
    "title":"Whose Truth Is It Anyway?
    - Al Jazeera English",
    "url":"https://www.aljazeera.com/
    programmes/whose-truth-is-it-anyway
    /truth.html"},
  {
    "title":"Preauricular lymph nodes:
    Causes of swelling",
    "url":"https://www.medicalnewstoday
    .com/articles/325947"},
  {
    "title":"List of hematologic
    conditions - Wikipedia",
    "url":"https://en.wikipedia.org/
    wiki/List_of_hematologic_
    conditions"},
  {
    "title":"Plasmodium vivax -
    Wikipedia",
    "url":"https://en.wikipedia.org/
    wiki/Plasmodium_vivax"},
  {
    "title":"What treatments are there
    for the novel coronavirus
    (COVID-19)?",
    "url":"https://www.medicalnewstoday
    .com/articles/coronavirus-
    treatment"},
  {
    "title":"Taste - Wikipedia",
    "url":"https://en.wikipedia.org/
    wiki/Taste"},
  {
    "title":"Risks to Switching Diabetes
    Meds? 8 Questions Answered by an Expert",
    "url":"https://www.healthline.com/health/
    type-2-diabetes/ask-the-pharmacist"},
  {
    "title":"Sutent: Side effects, dosage,
    uses, and more",
    "url":"https://www.medicalnewstoday.com/
    articles/sutent"},
  {
    "title":"Bell’s palsy - Wikipedia",
    "url":"https://en.wikipedia.org/wiki/
    Bell%27s_palsy"},
  {
    "title":"Coronaviruses: Symptoms,
    treatments, and variants",
    "url":"https://www.medicalnewstoday.com/
    articles/256521"},
  {
    "title":"This is Europe - Al Jazeera
    English",
    "url":"https://www.aljazeera.com/
    programmes/this-is-europe/"}]
}}
```

IV. Experiments

Experiments were conducted on MacBook Pro 13-inch 2019 model which has macOS Catalina version 10.15.4, 8 GB 2133 MHz LPDDR3 RAM, and 2.4 GHz Quad-core Intel Core i5 processor. We have used Reciprocal Rank and Precision at k measure to evaluate our solution.

a) MRR: Mean Reciprocal Rank is a measure to evaluate systems that return a ranked list of answers to queries. For a single query, the reciprocal rank is 1/rank where rank is the position of the highest-ranked answer (1,2,3,. . . ,N for N answers returned in a query). If no correct answer was returned in the query, then the reciprocal rank is 0. For multiple queries Q, the Mean Reciprocal Rank is the mean of the Q reciprocal ranks.

b) P@k: Precision at k is the proportion of recommended items in the top-k set that are relevant. Suppose that my precision at 10 in a top-10 recommendation problem is 80%. This means that 80% of the recommendation I make are relevant to the user. Precision@k = (# of recommended items @k that are relevant) / (# of recommended items @k)

We have crawled about 32k documents. Out of which more than 30k documents are html pages which are relevant for our text based query. There are very few documents crawled for audio, video and other application type which is not considered useful in our scenario. Thus we can say majority of documents is of good quality.

We have shown results in Table IV. Most of our queries for symptoms returned correct disease result at first position thus many queries have Reciprocal Rank as 1. Given the small size of our crawled database we have done a pretty good job by displaying correct result in one of the First Top-10 positions though we could not fetch correct result for some queries. However we have configured our UI to display 25 results per page because some of the queries fail to display correct disease for a symptom in Top-10 list but can do so in Top-20 list. Thus it becomes important to display at least Top-20 results. We also show how we correctly diagnose some variations on the symptoms as we have shown in the table in all 3 cases for Covid-19 we successfully retrieve the correct result in the first position.

TABLE III: Types and Count of Indexed Documents

Document Type	Count
text / html	30131
application / xhtml + xml	1251
application / pdf	573
image / svg + xml	186
image / jpeg	104
application / rss + xml	87
image / png	22
text / plain	21
application / xml	20
audio / vorbis	8
image / gif	8
application / msword	4
application / vnd.openxmlformats-officedocument.presentationml.presentation	4
image / vnd.microsoft.icon	2
application / atom + xml	1
application / vnd.openxmlformats-officedocument.wordprocessingml.document	1
audio / mpeg	1
video / mp4	1

V. Source Code

Source code for MediCrawl implementation of is publicly available at https://github.com/devharsh/MediCrawl.

VI. Conclusions

In this project we designed and implemented "MediCrawl" - A web search engine for medical diagnosis for common and rare diseases in children and adults which provide free text search capabilities. We have used Free and Open source tools to build easy to implement and portable solution. We have successfully implemented a solution where any user disregarding their medical expertise can search for any arbitrary symptoms and get a result of possible diseases.

In this process we faced following challenges and solved them:

- Establishing ground truth for query retrieval evaluation
  - For symptoms of common diseases:
    - ∗ https://www.advocatehealth.com/assets/documents/subsites/condell/ems/emsce/jan_2006_table_communicablediseases.pdf
    - ∗ https://www.healthed.govt.nz/resource-table/table-infectious-diseases
  - For symptoms of rare diseases:
    - ∗ https://www.sciencedirect.com/science/article/abs/pii/S1386505613000166?via%3Dihub
- Building own database of inverted index - we crawled and index more than 30,000 documents based on the seed list of following websites:
  1) https://www.omim.org/
  2) https://rarediseases.org/rare-diseases/
  3) https://www.seattlechildrens.org/conditions/a-z/
  4) https://kidshealth.org/en/parents/
  5) https://www.medicalnewstoday.com/
  6) https://my.clevelandclinic.org/health/diseases
  7) https://www.orpha.net/consor/cgi-bin/Disease_Search.php
  8) https://www.orpha.net/consor/cgi-bin/index.php
  9) https://www.healthline.com/
  10) https://www.health.harvard.edu/
  11) https://www.ncbi.nlm.nih.gov/books
  12) https://www.ncbi.nlm.nih.gov/mesh
  13) https://www.ncbi.nlm.nih.gov/pmc/
  14) https://www.nhsinform.scot/illnesses-and-conditions/a-to-z
  15) https://www.webmd.com/a-to-z-guides/common-topics
  16) https://www.webmd.com/a-to-z-guides/health-topics
  17) https://www.mayoclinic.org/diseases-conditions/
  18) https://www.cdc.gov/diseasesconditions/
  19) https://en.wikipedia.org/wiki/Category:Rare_infectious_diseases
  20) https://en.wikipedia.org/wiki/Category:Lists_of_diseases
- Retrieve from large set of documents efficiently - we have used Solr LTR based Learning Model which is based on Nutch default BM25 ranking to retrieve documents efficiently within a few milliseconds.
- Ranking and Searching issues with long tailed query - using stemming of stop words and based on term frequency in a document.
- Managing large (binary) files - since we have created our own database it becomes difficult to manage and version control large files, we are using Git-LFS for this.

VII. Future Work

Future work could include improving this existing solution. Current database can be expanded to incorporate more diseases which have not been covered. Also we can add more information like medicines, treatments and remedies for a disease.

Another approach could be to use a deep learning solution instead of an information retrieval system like we have in this scenario. A deep learning system may take input parameters as token of symptoms with respect to different parameters like age, sex, location and severity of each symptom and generate a probabilistic list of diseases as output for a given query.

References

[1] "Mayo clinic: A search engine for common diseases,"

[2] R. Dragusin, P. Petcu, C. Lioma, B. Larsen, H. L. Jørgensen, I. J. Cox, L. K. Hansen, P. Ingwersen, and O. Winther, "Findzebra: A search engine for rare diseases," *International Journal of Medical Informatics*, vol. 82, no. 6, p. 528–538, 2013.

[3] "Db-engines ranking of search engines - 20 systems in ranking, april 2020."

[4] "Nutch search engine overview."

[5] "Introduction to lucene solr."

[6] "Learning to rank — apache solr reference guide 8.5."

TABLE IV: Query Performance Results

Index	Symptoms	Disease	RR	P@10	P@20
1	fever,lack of taste	Covid-19	1.00	0.40	0.25
2	shortness of breath lack of taste	Covid-19	1.00	0.30	0.45
3	shortness of breath, lack of taste, fever	Covid-19	1.00	0.50	0.45
4	Fever and spots with a blister on top of each spot	Chickenpox (Varicella)	0.13	0.20	0.15
5	Rash in mouth, hands (palms and fingers), and feet (soles); fever; loss of appetite; may be asymptomatic	Hand, foot and mouth Disease (Coxsackie Virus and Enterovirus Diseases)	1.00	0.20	0.10
6	Itchy scalp, especially behind ears. Occasionally scalp infections that require treatment may develop	Head lice (Nits)	-	-	-
7	fever, runny nose, white spots in the back of the mouth and a rash Runny nose and eyes, cough and fever, followed a few days later by a rash Cough, runny nose, conjunctivitis, fever, rash that starts at head and spreads down and out on body; sore throat; may have Koplik's spots	Measles (Rubeola virus)	1.00	0.20	0.15
8	Flat, ring-shaped rash Skin: red circular patches with raised edges and central clear area; cracked peeling skin between toes; Scalp: redness, patchy scaly areas with/without hair loss Can occur in multiple sites on the body	Ringworm (e.g. tinea corporis, tinea capitis)	1.00	0.10	0.05
9	Fever, swollen neck glands and a rash on the face, scalp and body	Rubella (German Measles)	0.10	0.10	0.10
10	Red bumps commonly found in skin folds; burrows appear as tiny whitish or gray lines on skin surface; intense itching, especially at night	Scabies	1.00	0.30	0.15
11	Blisters on the body which burst and turn into scabby sores	School sores (Impetigo) Staphylococcus or Streptococcus bacteria	-	-	-
12	Red cheeks and lace-like rash on body Redness of the cheeks and body, "slapped cheek" rash. May have mild fever, runny nose, headache Rash may come and go for weeks	Slapped cheek (Human parvovirus infection) Fifth Disease	1.00	0.10	0.05
13	mild to severe diarrhea, bloody diarrhea, stomach pain, cramps, nausea and/or vomiting, fever, headache, and muscle pain	Campylobacter / Salmonella	-	-	-
14	Nausea, stomach pains, general sickness. Jaundice a few days later Fever, loss of appetite, nausea, abdominal discomfort and weakness followed by jaundice. Many unrecognized mild cases without jaundice occur, especially in children	Hepatitis A	1.00	0.20	0.15
15	Nausea, vomiting, watery diarrhea, abdominal pain, possibly low-grade fever, chills, headache Duration of symptoms usually 12-72 hours	Norovirus	1.00	0.10	0.10
16	Sudden onset of fever with cough, sore throat, muscular aches and a headache Sudden onset of fever, chills, headache, malaise, body aches, and nonproductive cough	Influenza and Influenza-like illness (ILI)	0.25	0.50	0.30
17	Headache, vomiting, sore throat. An untreated sore throat could lead to Rheumatic fever	Streptococcal sore throat (Rheumatic fever)	1.00	0.20	0.15
18	causes coughing spells that can last for weeks Runny nose, persistent cough followed by "whoop", vomiting or breathlessness Initially cold-like symptoms, later cough; may have inspiratory whoop, post-tussive vomiting	Whooping cough (Pertussis)	1.00	0.30	0.15

19	Irritation and redness of eye. Sometimes there is a discharge Red eyes, usually with some discharge or crusting around eyes; may be itchy, sensitive to light, or watery Bacterial: may have yellow/greenish discharge; may affect one or both eyes Allergic and chemical conjunctivitis usually affects both eyes	Conjunctivitis, Bacterial or Viral (Pink eye)	0.33	0.10	0.05
20	swelling of salivary glands, usually the parotid glands	Mumps	0.50	0.30	0.20
21	headache, malaise, fever, and swollen salivary glands Pain in jaw, then swelling in front of ear and fever	Mumps	0.50	0.30	0.20
22	infection that most often attacks the lungs, but in infants and young children, affects other organs like the brain Fever, fatigue, weight loss, cough (lasting 3+ weeks), night sweats, loss of appetite	Tuberculosis	1.00	0.20	0.10
23	liver infection	Hepatitis B	0.25	0.10	0.10
24	breathing muscles are paralyzed	Poliovirus	1.00	0.40	0.30
25	infects the throat and tonsils, making it hard for children to breathe and swallow	Diphtheria	1.00	0.30	0.15
26	painful muscle contractions	Tetanus	-	-	-
27	pneumonia, meningitis and other severe infections almost exclusively in children under 5 years old	Haemophilus influenza type b (Hib)	1.00	0.10	0.30
28	genital warts in both men and women, as well as cancer on other parts of the body	Human papillomavirus (HPV)	1.00	0.30	0.25
29	Fever, runny nose, cough. May have wheezing	Bronchiolitis-RSV	0.25	0.20	0.15
30	Fever, sore throat with pus spots on tonsils, tender swollen glands	Strep throat/Scarlet Fever	0.50	0.20	0.10
31	Profuse, watery diarrhea, sometimes with blood and/or mucus, abdominal pain, fever, vomiting	E. coli (Escherichia coli) infections	0.06	0.00	0.05
32	Abdominal pain, diarrhea (possibly bloody), fever, nausea, vomiting, dehydration	Salmonellosis	0.20	0.10	0.05
33	Diarrhea, which can be profuse and watery, preceded by loss of appetite, vomiting, abdominal pain; asymptomatic cases can spread the infection to others; symptoms can come and go for up to 30 days	Cryptosporidiosis	0.08	0.00	0.10
34	Often asymptomatic, but itching around the anus is a common symptom	Pinworms (Enterobius vermicularis)	1.00	0.20	0.10
35	Primary infections may have no symptoms; may have fever or malaise; may or may not have rash, vesicular lesions, or ulcers at site, “fever blister”/cold sore HSV 1 and HSV2 lesions can appear on other parts of the body.	Herpes Simplex (cold sores, skin lesions) HSV1 (cold sores) HSV2 (genital lesions)	0.13	0.10	0.05
36	Skin lesions such as furuncles (abscessed hair follicles or “boils”), carbuncles (coalesced masses of furuncles), and abscesses; may have purulence (pus), yellow/white central point, redness	MRSA (Methicillin-resistant Staph aureus) skin infections	1.00	0.10	0.05
37	Small flesh colored bumps on the skin that may have a tiny indented center	Molluscum contagiosum	1.00	0.10	0.05
38	Painful rash that develops typically on one side of the body; may have fever, headache, chills, nausea	Shingles/Zoster	0.17	0.10	0.15
39	Fever, cough, sore throat, muscle aches, eye infections (conjunctivitis), acute respiratory distress, viral pneumonia.	Avian or Bird Flu	0.11	0.10	0.20
40	fever, headache, body aches, occ rash on trunk, swollen lymph glands	West Niles Virus	-	-	-
41	12 days after exposure get fever, headache, muscle aches, backache, swollen lymph nodes, tired	Monkeypox	0.14	0.10	0.05
42	Chills, high fever, dyspnea, pleuritic chest pain worsened by deep inspiration, cough, crackles & wheezes heard on breath sounds	Pneumonia	0.25	0.20	0.15

43	Mono-like syndrome, fatigue, fever, sore throat, lymphadenopathy, splenomegaly, rash, diarrhea. Skin lesions (Kaposi's sarcoma); opportunistic infections (Pneumocystic carinii pneumonia, Tb)	HIV	-	-	-
44	delusions, hallucinations, disorganized speech, lack of motivation or emotion	Schizophrenia	1.00	0.70	0.40
45	Boy, normal birth, deformity of both big toes (missing joint), quick development of bone tumor near spine and osteogenesis at biopsy	Fibrodysplasia ossificans progressiva	0.20	0.10	0.10
46	25 year old, woman, conjunctival hyperaemia, interstitial keratitis, moderate bilateral sensorineural hearing loss, tinnitus, dizziness, nausea and vertigo	Cogan's syndrome	1.00	0.10	0.05
47	11 year old, boy, severe psychomotor retardation, seizures, strabismus, inverted nipples, dilated cardiomyopathy, hypotonia, wheelchair-bound	CDG (Congenital Disorders of Glycosylation) syndrome type Ic. (Synonyms: Carbohydrate deficient glycoprotein syndrome type Ic, Congenital disorder of glycosylation type 1c (or Ic))	1.00	0.10	0.05
48	11 year old, girl, intermittent abdominal pain, mild dorsal scoliosis, low serum phosphate/hypophosphatemia, hypercalcuria, elevated serum 1,25 dihydroxyvitamin D	Hypophosphatemic rickets with hypercalciuria	0.10	0.10	0.05
49	46 year old, female, ptosis, acanthocytosis, history of diarrhea, ataxia, paresthesia	Abetalipoproteinemia (ABL). (Synonyms: Bassen-Kornzweig disease, Homozygous familial hypobetalipoproteinemia (HoFHBL))	1.00	0.10	0.05
50	16 year old, girl, persistent diarrhea, acanthocytosis, mild dysarthria, reduced muscle bulk, bilateral proximal muscle weakness, absent deep-tendon reflexes, upgoing plantar reflexes, reduced sensitivity to light, dysdiadochokinesia	Abetalipoproteinemia (ABL). (Synonyms: Bassen-Kornzweig disease, Homozygous familial hypobetalipoproteinemia (HoFHBL))	1.00	0.10	0.05
51	Acute Aortic regurgitation, depression, abscess	Infective endocarditis	0.17	0.10	0.05
52	fever, anterior mediastinal mass and central necrosis	Lymphoma	0.11	0.10	0.05
53	multiple spinal tumours, skin tumours	Neurofibromatosis type 1	0.08	0.00	0.05
54	bullous skin conditions, respiratory failure, carbamazepine	Toxic Epidermal Necrolysis Syndrome (TENS)	1.00	0.10	0.05
55	polyps, telangectasia, epistaxis, anemia	MADH4 mutation (HTT + juvenile polyposis)	-	-	-
56	buttock rash, renal failure, edema	Cryoglobulinaemia	-	-	-
57	renal transplant, fever, cat, lymphadenopathy	Cat scratch disease	-	-	-
58	myopathy, neoplasia, dysphagia, rash, periorbital swelling	Dermatomyositis secondary to NHL	1.00	0.20	0.10