

ECE:5995 Modern Databases – Fall 2020

Homework 5 – Indexing in MongoDB

Due Tuesday Oct 27th, 2020 11:59pm.

For this homework we will use the zip code collection and a new collection for paper abstracts. The homework has two parts. In the first part we will create a two-dimensional index and in the second one we will create text indexes.

Part 1. Spatial indexing in MongoDB.

MongoDB geospatial tutorial link:

<https://docs.mongodb.com/manual/tutorial/geospatial-tutorial/>

1. Add a 2dsphere spatial index for the location of the zipcodes collection using the `createIndex` command. What is the size of the loc index in KBs? You can use the `stats` command over the collection to find out the index size.
2. Write a query using `find` to retrieve the zip code with location `[-91.511192, 41.654899]`.
3. Write a query using `find` to retrieve all the zip codes within 10 miles of a circle centered at location `[-91.511192, 41.654899]` using the `$geoWithin` and `$centerSphere` commands.
4. Write a query using `find` to retrieve all the zip codes contained in the rectangle with left bottom corner at `[-91.6143, 41.6623]` and upper right corner `[-91.123750, 41.873890]` using the `$geoWithin` and `$box` commands.
5. Write a query using `find` to retrieve the zipcodes within a `minDistance` of 1,913.43 km (0.3 radians) and a `maxDistance` of 1,914.71 km (.3002 radians) from location `[-73.9667, 40.75]` using the `$nearSphere` command.

For these queries, locations are specified as longitude, latitude pairs which is the order expected by MongoDB (as opposed to latitude, longitude).

Part 2. Text (inverted) indexes in MongoDB.

<https://docs.mongodb.com/manual/text-search/>

1. Download the `abstracts.csv` file from ICON and import it into an `abstracts` collection in your `mydb` database using the `mongoimport` and specifying `--type csv` and `--headerline` parameters. The `abstracts` collection should have 398 documents containing title, authors, groups, keywords, topics, and abstract. All text value attributes. You can examine the schema of the documents using the `findOne` command.
2. Create a text index on the title and authors. You can provide a name (later we will drop this index). How big is the index in KBs?

3. Write the following queries using the find command (to use the text index in a query you need to use the \$text and \$search commands). For all the queries, only display the _id and the title of the documents.
 - a. Search for the documents with the keywords “gamble” or “approximation” or “gaspers” in the text index.
 - b. Search for the documents with the phrase “security games”
4. MongoDB only supports one text index per collection. Drop the text index for the title and create another text index for all the attributes in the collection. You can specify each key name as parameters or use the \$** to indicate all the attributes. How does the size of this index compare to the size of the collection?
5. When using a text index, you can sort the results by an internal score computed by mongo: {score: {\$meta: 'textScore'}}. Retrieve the top 10 documents with the word “approximation” or “probability” in the text index, display the title and the score.

Submission:

Submit your homework as a single file to the dropbox for Homework 5 in ICON by **Tuesday October 27th 11:59pm.**