Christopher Rech - 124000677

Andrew Allen - 124008219

Task 2 Report: GroupMe Search Bot

The second component of the GroupMe Search Bot is the search engine. A functional search engine was created in this portion of the project, with code to insert and query the data from the previous task. The search engine was created through Elasticsearch, which is based on the Lucene library. Elasticsearch is very similar to Solr, in that it provides all the same basic functionalities of indexing and searching, as well as being based on the same underlying platform, but our group found that it would be easier to interface with Elasticsearch than Solr.

Elasticsearch operates through HTTP requests, just like Solr, but our group simplified this process through the Elasticsearch Python library, which allows for basic requests to be made with simple and straightforward commands. As the prior work on the project had already been done in Python, the Elasticsearch Python library was a natural means of handling search indexing and querying.

The code to insert messages and execute search queries is shown in the provided video. The “push_from_file” function calls upon the Elasticsearch library in order to perform a bulk indexing of all of the messages in a file. In Elasticsearch, messages are indexed by index and type, where the type will always be “message” due to the nature of the data being stored. The index field is very relevant, however, because it may be important for users to only search for messages within one of their groups on the GroupMe app. Thus, when a search is performed on a group, the index will filter out results from other groups. In the “source” field, the sender, text, and timestamp of the message are stored. Thus, all of the information queried in the previous part of the project is retained and stored in the second part.

The “searchByKeyword” function also calls upon the Elasticsearch library, this time to perform a search. The information passed to the function is the filename and the keyword to be searched. The video example shows the search of two different keywords on one of the files which had been pushed into Elasticsearch. The search is completed successfully, in that the results shown all contain the keyword. Manual testing was conducted to check if any results were mistakenly omitted, and they were not. Thus, based on a few tests, the search had perfect accuracy and precision, with no false positives or false negatives. This is due both to the robustness of Elasticsearch and Lucene, as well as the relative simplicity of the search. For more complicated search terms, the same accuracy and precision are not guaranteed.

In the third and final component of the project, an interface will be constructed in order to query the functions demonstrated in this video. When a user grants permission to the GroupMe Search Bot, the content of their group will be indexed, as demonstrated in the video. When a search is conducted, the functionality to execute the search will be called upon, as also demonstrated in the video. Lastly, the results will be displayed to the user, rather than just to the console.

The end result of the next step will be that users will be able to execute searches in the GroupMe app, by means of our project. This is useful because the GroupMe app, despite having millions of users, does not actually have a built-in functionality to search through the messages of a group. Thus, the introduction of search functionality is expected to be highly appreciated by the prospective users of the app.