

Christopher Rech - 124000677

Andrew Allen - 124008219

### Project Report: GroupMe Search Bot

The GroupMe Search Bot is designed to solve a major issue in the popular GroupMe messaging app: the lack of a built-in search functionality, as is present in other popular messaging apps. GroupMe is used by nearly every student at Texas A&M, but many of them lament its lack of functionality compared to other platforms. The GroupMe Search Bot will handle search requests by GroupMe users, thus rectifying this issue and providing value to its users. The GroupMe Search Bot will incorporate information retrieval methods through its indexing of messages and querying of search terms.

The past work relating to information retrieval which has been done on GroupMe is limited, despite the popularity of the app and the functionality of its API. There are various existing projects which perform analytics on GroupMe, such as returning the members of the group with the most likes. Additionally, there exist a limit number of clients which offer search functionality for GroupMe. However, the problem with these applications is that they require total access to be granted by the user. This issue is resultant from the way the GroupMe API is set up. In order to retrieve a group's messages through a normal application, the application requires a user's access token. This single access token not only gives access to messages for a group, but also the messages for every single group and the ability to send messages. Thus, users are hesitant to use applications which require their access token to be given out, because a malicious agent with their access token could cause considerable privacy violations and reputational damage to the user. The GroupMe Search Bot resolves these security concerns by never requiring a user's access token. Instead of functioning as an application, it functions as a

GroupMe user, with its messages being automated by a program using its own access token. The bot retrieves data not by requesting a user's access token as is done in other applications, but rather by being added to the group on which the search is to be performed. This way, the user never has to give away any information or access other than the baseline requirements in order to perform the search.

Group statistics will vary by user. The following table presents statistics for the number of messages per group from the 100 most recent GroupMe chats of one of our team members.

Statistic	Number of Messages in Group
Minimum	1
10 <sup>th</sup> Percentile	29
25 <sup>th</sup> Percentile	83
Median	173
75 <sup>th</sup> Percentile	532
90 <sup>th</sup> Percentile	2255
Max	34680
Mean	1503

As one can tell from the above table, the distribution of messages per group is highly skewed right, meaning that a small handful of groups contain the vast majority of messages. Although there are a total of 150 thousand messages available, only 34 thousand of these are provided, due to privacy concerns from the members of some of the groups. The search functionality is expected to be most useful in the 10% of groups containing a very high number of messages.

The main problem encountered was in creating valid queries for the GroupMe API, for which there are very few examples of the various functions in its documentation. This problem was resolved by using a wrapper called Groupy, which simplifies the process of making requests to the GroupMe API. The other main problem encountered was that the GroupMe API frequently returns a request timeout error, or HTTP 408. This problem will be resolved in the application by repeatedly making queries until a valid response is returned.