**Topic Modeling App Documentation**

Mina Narayanan

[minamjn@gmail.com](mailto:minamjn@gmail.com)

Note that every functionality for the app has not yet been built. For example, the user must always choose to upload files when prompted by the application. If the user chooses not to upload a file when prompted, the application will not finish running. Other functionalities that need to be built are described in the comments labeled “TODO” in the server.R, helpers.R, and ui.R files.

1. The user first uploads a grant application file. This file must be downloaded from QVR and contain the columns located in the awarded-grants-NINR-2019.xlsx file on Github.
2. The user is then prompted to upload a stopwords file. The file must contain words that are formatted in a single column. This format is demonstrated in the example\_stopwords.xlsx file.
3. The user then has the option to upload LDA parameters. This file must be formatted the same way as the LDA\_Opt\_Param\_NINR19.csv file. This file is created by running the LDA optimization algorithm which is not yet encoded in the app, but is coded into the /R/TopicModeling-NINR.Rmd file of any branch of the TopicModeling-NINR Github repository.
4. The user can adjust the value of lambda on a scale of 0 to 1. When lambda is closer to 0, the words of each topic group are those that are most predictive of the topic. In other words, setting lambda close to 0 upscales words that are not prevalent across the entire dataset but are unique to a topic. When lambda is set closer to 1, the words of each topic group are those that are more closely associated with a topic.
5. The words “Processing has started” become grayed out as the LDA algorithm creates topic groups.
6. When processing is complete, the words “Image files and topic groups have been saved to the app directory” appear on the user interface. An interactive LDA plot is embedded in the interface, which shows an intertopic distance map and the top-30 most salient terms of each topic. The percentage breakdown of topics by grant and the topics themselves are saved to .csv files in the app directory. Images of the intersections of topics within grants, scatter plots of topics by program officer, the distribution of topics by program officer, and the frequency of topics in portfolios are also saved to the app directory. Currently, the user is not able to customize the names of these files or where they are saved.