**Code and Data required to create the BRAINWORKS application**

5/1/2023

This directory contains the code and data needed to create the BRAINWORKS application as of February 22nd, 2023. In this document we provide an overview of the contents within this directory and considerations for how to use it. The intended audience of this documentation is a software developer. However, we also include an overview of the development activities in *documents/project\_description\_2022.pptx*

**Data**

In the directory entitled */data* are all data (in CSV format) that were created and used during the BRAINWORKS project. The directory consists of files which obey the following naming convention:

* *<table-name>manifest:* Where <table-name> is the name of the database table. This file contains a list of all files that are segments of the same table; the manifest is used by an AWS RedShift database to load the table from the proper segments.
* *<table-name>000<N>\_part\_00:* Where <table-name> is the name of the database table, and where <N> signifies the segment of the table contained within the file.
* *MR<table\_name>.RRF:* These files contain UMLS Metathesaurus tables that are necessary for the concept analysis performed by the application.

To use these files:

1. Upload all files to an S3 bucket within AWS.
2. Load each table (spread out over multiple files), into a RedShift database following the documentation available here: <https://docs.aws.amazon.com/redshift/latest/dg/copy-parameters-data-source-s3.html>

Note: These steps were performed for a RedShift cluster within the NIH environment, which serves as the backend for the currently deployed website; the cluster was functional as of February 22nd, 2023.

**Code**

In the directory entitled */code* are three sub-directories; each sub-directory contains a README.md file with additional details on the sub-directory’s contents:

1. /*data\_collection\_and\_preperation\_pipeline* contains all code necessary to collect, clean, and format all data used by the BRAINWORKS application for storage within the AWS RedShift database.
2. */graph\_visualization\_API* contains all code from the external API used to render graph visualizations within the platform; this was not part of the formal BRAINWORKS deliverable.
3. */web\_application* contains all code necessary to deploy the front-end web application hosting the BRAINWORKS website.

**Documentation**

In the directory entitled */documentation* are three files:

1. *system\_diagram.png:* This file depicts the system infrastructure as it was deployed within the NIH environment as of February 22nd, 2023. Note that the points labelled Developer Control will require maintenance or administration by an internal tech lead.
2. *project\_description\_2022.pptx:* This PowerPoint presentation details the functionality and features of the BRAINWORKS application developed during 2022.
3. *proposed\_post\_project\_development\_activities.xlsx*: This Excel spreadsheet contains a list of suggested prospective development activities.