1. Installing the Program

Note: These instructions assume that you are in a Linux environment (you should be...)

1.1. Clone repo and change into it

```
git clone https://github.com/jramsdell/cs753_team2_assignment_3.git
cd cs753_team2_assignment3
```

1.2. Compile program

You can either use the precompiled program located in bin/program.jar or you can compile it yourself. Compilation requires maven (skip the first line if it's already installed) and that you be in the project directory:

```
sudo apt-get install maven
./compile.sh
```

This will create a new jar file located at: target/team2_3-1.0-SNAPSHOT-jar-with-dependencies.jar

2. Indexing

Note: For brevity's sake, we will be referring to the precompiled jar for the following commands. You can replace this with the one compiled in target if you like.

While in the project directory, call:

```
java -jar bin/program.jar index PARAGRAPHS
```

Where PARAGRAPHS is the location of the paragraphs .cbor file that you wish to index with this command. The output of this command is a new directory (created in the project directory) called paragraphs/

3. Generating Runfiles

While in the project directory, call:

```
java -jar bin/program.jar search paragraphs/ OUTLINES
```

Where OUTLINES is the path to the outlines .cbor file. This will create a new directory called "results/" and inside of this directory will be the runfiles for BM25 (runs with "standard" in their name) and each of our TFIDF variants. The runfiles are prefixed by page- or section- to indicate which type of query they were (use article qrels on page runs, and hierarchical qrels on section runs).

4. Getting Correlations

To get the spearman results (correlations for each TFIDF run against the standard BM25 run), do the following:

```
java -jar bin/program.jar spearman RUNFILE_DIRECTORY QRELS
```

Where RUNFILE_DIRECTORY is the results directory created by running the **search** command in the previous section (so if you are in the project directory, you would probably be pointing to results/).
Calling this command will print a newline-delimited list of the correlations of each TFIDF variant against that of BM25. Note that the prefix (section- or page-) indicates what type of query was used to compare the variants with.