

Build Script (Team1)

Amith Ramanagar Chandrashekar Pooja Himanshu Oza
Rachel E Cates Ahmed M Alnazer

February 21, 2019

1 Dependency

Java version: Java8 or Higher

Build Tool: Maven

2 How to Run

Please perform the below steps

- `sudo git clone https://gitlab.cs.unh.edu/cs953-2019/cs953-team1.git`
- `sudo chmod 777 cs953-team1`
- `cd cs953-team1`
- `./install.sh`

Running the `./install.sh` creates all the runs files in the current working directory, some of the run files will be created in the result folder. we have the Benchmark-Y1 test as the query file in the install script.

Self executable jar can be created using the maven and the program can run independently without the install script. Please follow below steps

- `mvn clean compile`
- `mvn package`
- `java -jar cs953-team1-1.0-SNAPSHOT-jar-with-dependencies.jar sub-commands option`

2.1 Note

The below error can be ignored. This error is from the Nd4j.

SLF4J: Failed to load class "org.slf4j.impl.StaticLoggerBinder".
SLF4J: Defaulting to no-operation (NOP) logger implementation
SLF4J: See <http://www.slf4j.org/codes.html#StaticLoggerBinder> for further details.

A entity graph based is taking a while to run the program, it can be run inside the tmux as follows.

- `tmux new -s team1` (create new session)
- Run the `./install.sh` inside the tmux
- `ctrl+b+d` (Disconnect from the session)
- `tmux a -t myname` (To attach to the session)

3 Command Line Option

These are the command line options we have implemented.

Usage: `<main class> [command] [command options]`

Commands:

`index` Command to Index the Corpus

Usage: `index [options]`

Options:

* `-i, --corpus-file`

Corpus file to index

`-d, --dest-location`

Location to save the index file

Default: `/home/team1/cs953-team1/indexed_file`

`--entity-index`

Perform Entity Index

Default: `false`

`--help`

`--para-index`

Perform Paragraph Index

Default: `false`

`search` Command to search

Usage: `search [options]`

Options:

`--bias-fact`

```

    Bias factor to get the document representation
    Default: 1
-k, --candidate-set-val
    How many candidate set to retrieve using BM25
    Default: 100
--cosine-sim
    Rerank the document based on the cosine similarity between two
    strings
    Default: false
-bm25, --default-bm25
    Rerank the initial retrieved cluster using document similarity
    Default: false
--dice-sim
    Rerank the document based on the Sorensen Dice coefficient
    similarity between two strings
    Default: false
--entity-degree
    Rerank the initial retrieved document using entity degree
    Default: false
--entity-expand
    Rerank the initial retrieved document using expanded query
    Default: false
--entity-index
    Pass the index location of entity index
--entity-sim
    Rerank the initial retrieved document using entity abstract
    similarity
    Default: false
--help

* -i, --index-loc
    Indexed directory to search
--jaccard-sim
    Rerank the document based on the Jaccard similarity between two
    strings
    Default: false
--jaro-sim
    Rerank the document based on the Jaro Winkler similarity between
    two strings
    Default: false
* -q, --query-cbor
    Query file (CBOR file)
-qe, --query-expansion
    Rerank the document using Query expansion
    Default: false
--rerank

```

```

    Rerank the initial retrieved document using document similarity
    Default: false
--rerank-df
    Rerank the document based on the DF
    Default: false
--rerank-idf
    Rerank the document based on the IDF
    Default: false
-V, --verbose
    Print out some of the results into stdout
    Default: false
-dim, --word-dimension
    Dimension of the Word embeddings
    Default: 0
-we, --word-embedding
    Pass the word embedding file GloVe/ Word2Vec
-top
    specify the top number of selected entity to used in the Query
    expansion
    Default: 3

indexHamSpam      Command to create ham-spam index
Usage: indexHamSpam [options]
Options:
* -i, --corpus-file
    Index
-h, --dest-hamtrain
    Location to save the ham training data
    Default: /home/team1/cs953-team1/indexed_file
-s, --dest-spamtrain
    Location to save the spam training data
    Default: /home/team1/cs953-team1/indexed_file
-d, --dest-test
    Location to save the test data
    Default: /home/team1/cs953-team1/indexed_file
--help

-q, --new-qrels-file
    qrels file
    Default: /home/team1/cs953-team1/indexed_file

filter            Use the ham and spam data sets to detect and filter spam from
                  search results
Usage: filter [options]
Options:
* -h, --ham-index

```

```
        Location of the ham training data
--help

* -i, --index
    Location of the main index
* -s, --spam-index
    Location of the spam training data
* -t, --test-index
    Location of the test data

--help      Help Information
Usage: --help
```