

Baseline and Evaluation

Name: Kunpeng Xie, Xin Liu

1. Download Data

We chose the data of WikiQA which is from Microsoft Research WikiQA Corpus. Data link: <https://www.microsoft.com/en-us/download/details.aspx?id=52419>

2. Project Structure

BackEnd:

Core:

Index

Main Search Functions

Ranking Functions:

a. LanguageModel

b. TF_IDF_anc_apc

c. TF_IDF_bnn_bnn

d. TF_IDF_Inc_ltn

FrontEnd:

Html page: Input search queries and results.

3. Implement a first method

Our first method is main method. The main method first read data from "WikiQA.tsv" file and indexed the section of Question and DocumentTitle. Then we used DocumentTitle as query and put all queries into Lucene Search Engine to get all questions.

4. Install Instruction

Required Tool:

1. Maven3.5.4: To automatic loading up required dependencies, download from: <https://maven.apache.org/download.cgi>
2. Git: Version control, download from <https://git-scm.com/downloads>
3. IntelliJ IDEA: download from <https://www.jetbrains.com/idea/download/#section=mac>

Installation:

1. Maven installation guide <https://maven.apache.org/install.html>

2. Git installation: <https://git-scm.com/book/en/v2/Getting-Started-Installing-Git>
3. IntelliJ installation: <https://www.jetbrains.com/help/idea/install-and-set-up-product.html>

Run in command line:

1. clone **CS853Project** from github(<https://github.com/CodeFreddy/CS853Project>) and cd in to the directory which contains pom.xml and src folder.
2. run

```
mvn clean install
```

and

```
mvn compile
```

3. run

```
mvn exec:java -Dexec.mainClass="Main.main" -Dexec.args="indexPath filePath"
```

5. Evaluation

We are still writing evaluation method of precision@R and MAP. We will get evaluation score soon.