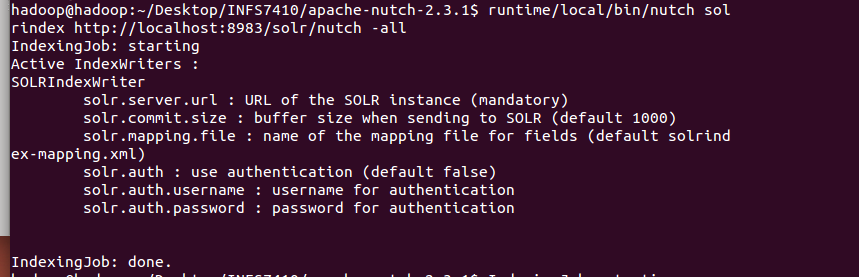
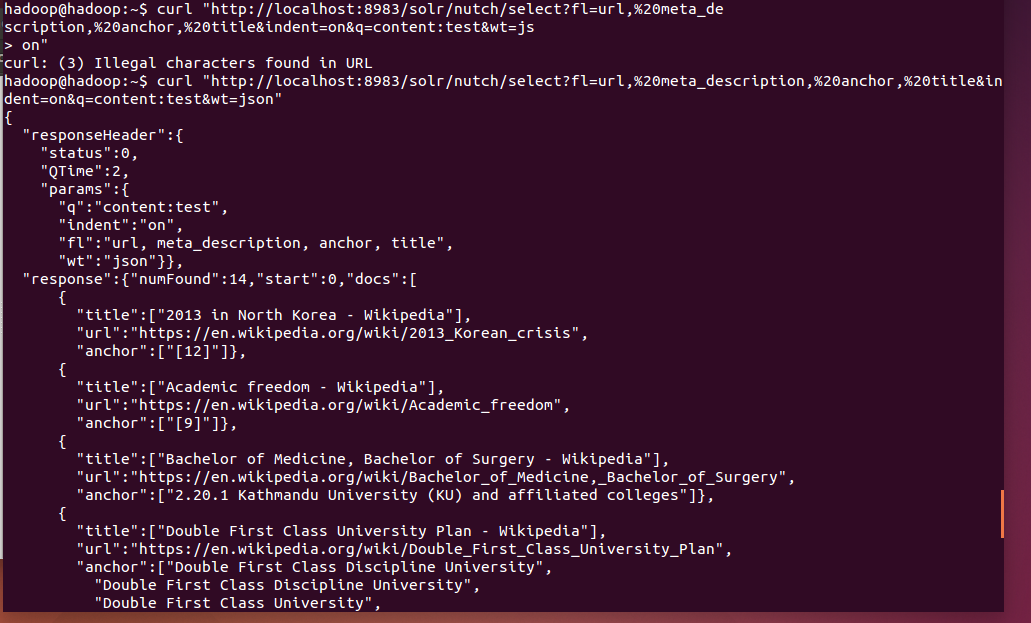
1. **Create an index for the crawled webpages using Solr**

Test indexes with curls:





1. **Create a simple search engine based on the crawled pages**

Query: The best university in C9

Default method (BM25 model):

Running time: 748 milliseconds

Precision = 0.2

Recall = 2/15

BM25 model with higher b value:

In this case, there is a need to increase the influence of document length in BM25.

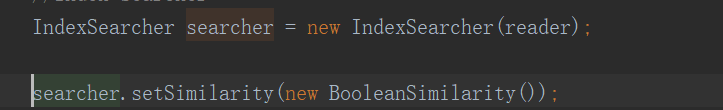




Running time: 572 milliseconds

The result is the same as the default BM25 model.

Boolean model:



Running time: 560 milliseconds

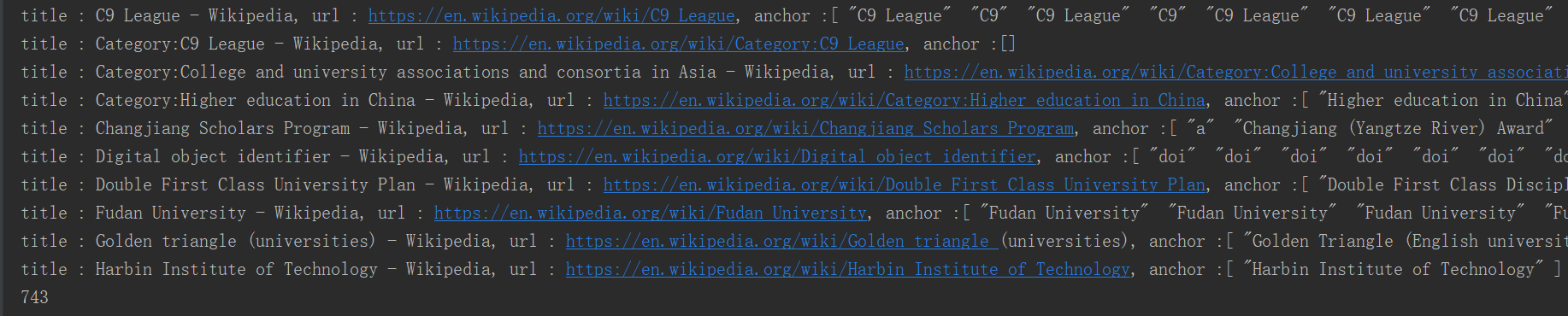
Precision = 0.7

Recall = 7/15

Language model based on the Jelinek-Mercer smoothing method:

Lower lambda value (good choice for shorter queries)





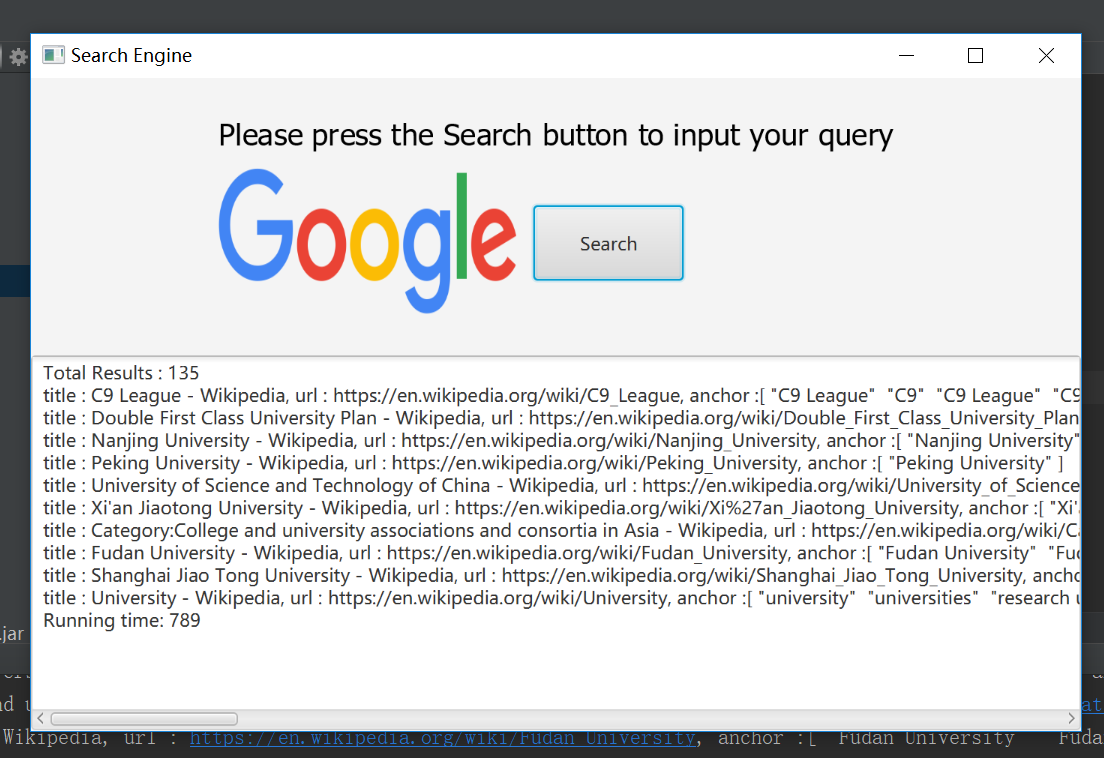
Running time: 743 milliseconds

Precision = 0.5

Recall = 5/15

According to the performances of different similarities mentioned above, the Boolean model was selected as the model of the search engine. Also, MultiFieldQueryParser was used to parse two fields (i.e. title weighs 3 and content weighs 1), compared with default QueryParser.

1. **Create graphic user interface for this search engine**

There are two interfaces for this project, one is JavaFx GUI and the other is Java ee web with highlighter.

