Assignment 4: Bigram Boolean Search

phuonglh@gmail.com

March 26, 2024

1 Problem

In this assignment, you will develop a program capable of searching for terms in unigram or bigrams, for example

- mr. tom AND jerry

- tom OR good jerry

Here, "mr. tom" or "good jerry" are two consecutive tokens (bigrams) in the original documents. You can also test your program on the Vietnamese corpus that you scraped in the last assignment.

2 Guide

You need to build the inverted index using both unigrams and bigrams. It is required that your program has a GUI to help build, load index and perform search. You are free to choose the GUI components and layout so that your interface is convenient for user.

3 Submission

If you develop your program in Scala, Python, or Java, submit your source file *YourFullName.scala*, *YourFullName.py*, or *YourFullName.java*, respectively; for example *NguyenPhuTrong.scala*. Your data file has the name *NguyenPhuTrong.json*. If you have multiple source files then put them in a directory and zip it to *YourFullName.zip* and submit before due date.