Twitter Keyword Search

Zhibin Huang, Mingdao Che, Ziyu Zhao

GitHub: https://github.com/mdche001/EC504_Twitter_Keyword_Search

Introduction:

With given tweets as database and a sentence (string) as input query, we want to get the most relevant tweet in our database. Just like tweet searching engine of google search engine. The search engine will take an input csv file as tweets database in which each line is one tweet.

Data Structure

We use inverted index to store our tweets. In computer science, an inverted index (also referred to as a postings file or inverted file) is a database index storing a mapping from content, such as words or numbers, to its locations in a table, or in a document or a set of documents (named in contrast to a forward index, which maps from documents to content). The purpose of an inverted index is to allow fast full-text searches, at a cost of increased processing when a document is added to the database [1]. By using inverted index, we could search through the tweets faster which only take O(1) time complexity for one key word. We process the tweets by the following steps:

Step 1: Split every line by space, store every unique word in lower case as the dictionary keys.
Step 2: For every key word, wo search through the tweets to know which tweets have such key word and store the result as a list value according to the key in the dictionary.

Algorithm:

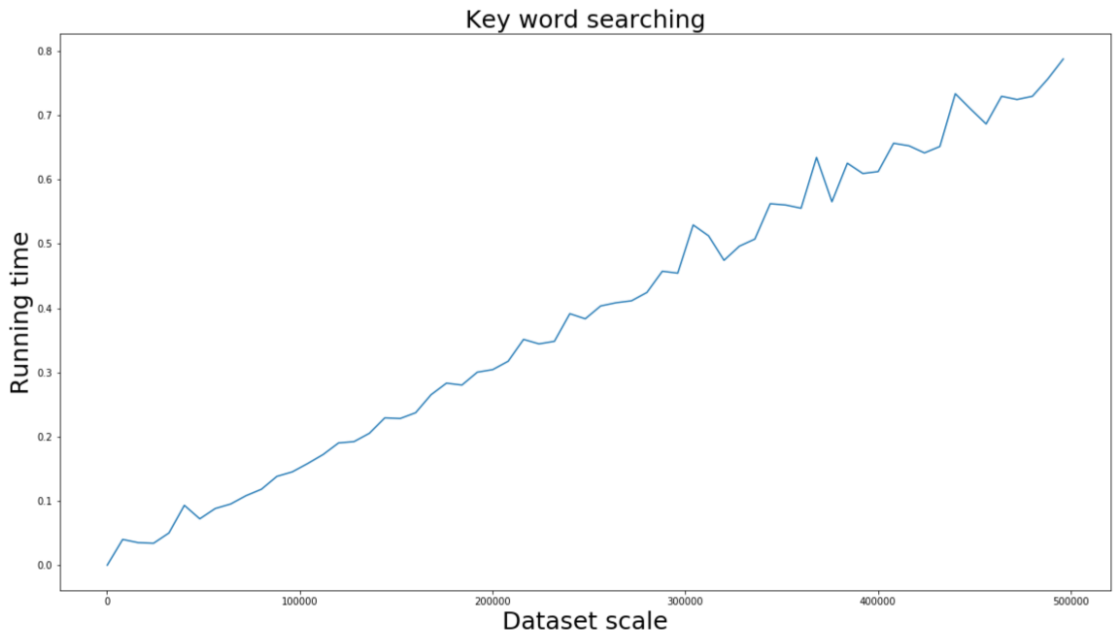
For each query with k key words in the string, it takes O(k) time for constructing a hash map. For i-th tweet with m words, time for calculating the number of key word in the tweet will be O(m). In fact, it's common case for different tweet with some same keys. So, assume there are total N tweets in the database, constructing the dictionary and finishing the k key words searching will take

$$O(k) + O\left(\sum_{i=1}^{N} m_i\right)$$

Finally, sorting the result with built-in function will take

$$O(N\log(N))$$

Time Complexity:



Technic Selection: Qt Designer + PyQt5 + Pycharm

Why Qt Designer：

Qt Designer is the Qt tool for designing and building graphical user interfaces (GUIs) with Qt Widgets. You can compose and customize your windows or dialogs in a what-you-see-is-what-you-get (WYSIWYG) manner and test them using different styles and resolutions [2].

Widgets and forms created with Qt Designer integrate seamlessly with programmed code, using Qt's signals and slots mechanism, so that you can easily assign behavior to graphical elements. All properties set in the Qt Designer can be changed dynamically within the code. Furthermore, features like widget promotion and custom plugins allow you to use your own components with Qt Designer.

The designer also provides people with the option of using Qt Quick for user interface design rather than widgets. It is a much easier way to write many kinds of applications. It enables a completely customizable appearance, touch-reactive elements, and smooth animated transitions, backed up by the power of OpenGL graphics acceleration.

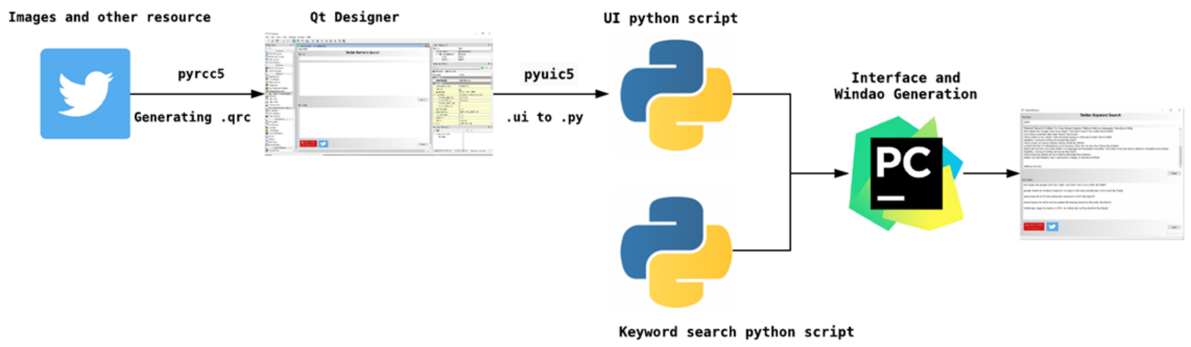
Why PyQt5:

PyQt is a set of Python v2 and v3 bindings for The Qt Company's Qt application framework and runs on all platforms supported by Qt including Windows, OS X, Linux, iOS and Android [3].

PyQt brings together the Qt C++ cross-platform application framework and the cross-platform interpreted language Python.
Qt is more than a GUI toolkit. It includes abstractions of network sockets, threads, Unicode, regular expressions, SQL databases, SVG, OpenGL, XML, a fully functional web browser, a help system, a multimedia framework, as well as a rich collection of GUI widgets.

Much of Python's power comes from its comprehensive set of extension modules providing a wide variety of functions including HTTP servers, XML parsers, database access, data compression tools and, of course, graphical user interfaces. Extension modules are usually implemented in either Python, C or C++. Using tools such as SIP it is relatively straight forward to create an extension module that encapsulates an existing C or C++ library. Used in this way, Python can then become the glue to create new applications from established libraries.

GUI diagram:



1) Material import:

The material is imported by .qrc file to the Qt Designer. The file of this project is in Figure.2.

```
<RCC>
<qresource>
  <file>Images/BU.png</file>
  <file>Images/twitter.png</file>
  <file>Images/background.png</file>
</qresource>
</RCC>
```

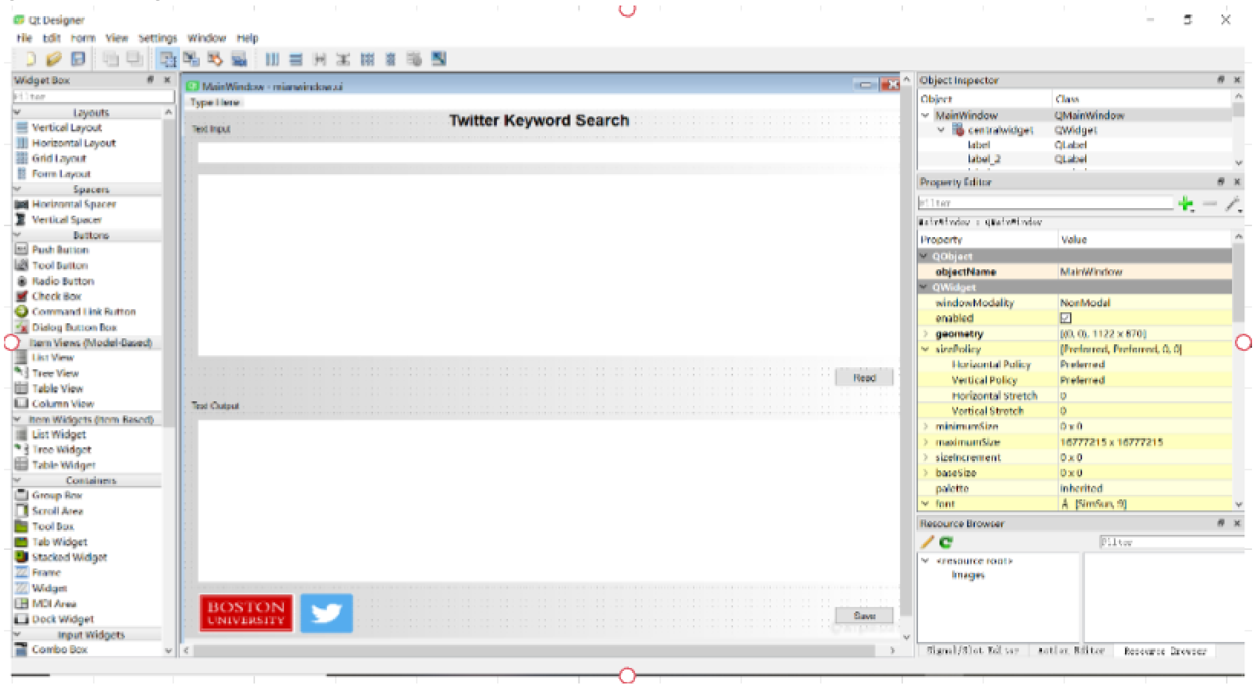
Figure.2 The .qrc file with 3 images

Then use qyrcc5 to generate the python resource file for python to call the resource.

```
pyrcc5 -o Images_rc.py Images.qrc
```

2) Layout Design:

In this Section, there are three main steps to design a simple main window GUI: choosing suitably components and Wdgets, arrange layouts and properties and generate signals and slots.



Object	Class
MainWindow	QMainWindow
centralwidget	QWidget
label	QLabel
label_2	QLabel

Property	Value
QObject	
objectName	MainWindow
QWidget	
windowModality	NonModal
enabled	✓
geometry	[(0, 0), 1122 x 870]
sizePolicy	[Preferred, Preferred, 0, 0]
Horizontal Policy	Preferred
Vertical Policy	Preferred
Horizontal Stretch	0
Vertical Stretch	0
minimumSize	0 x 0
maximumSize	16777215 x 16777215
sizeIncrement	0 x 0
baseSize	0 x 0
palette	Inherited
font	A [SimSun, 9]

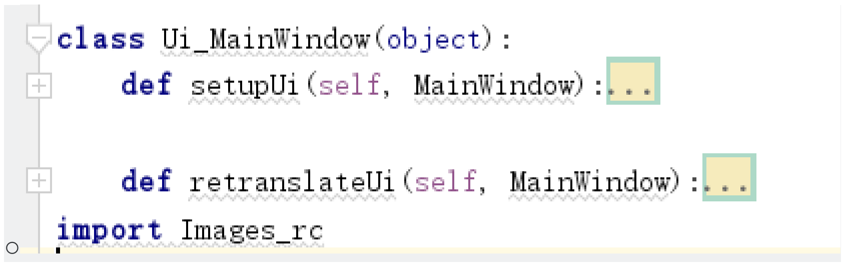
Figure.3 The Qt Designer

3) The Interface and MainWindow:

After layout design, you can use following command in terminal to transfer UI files to Python, which is based on the pyuic5

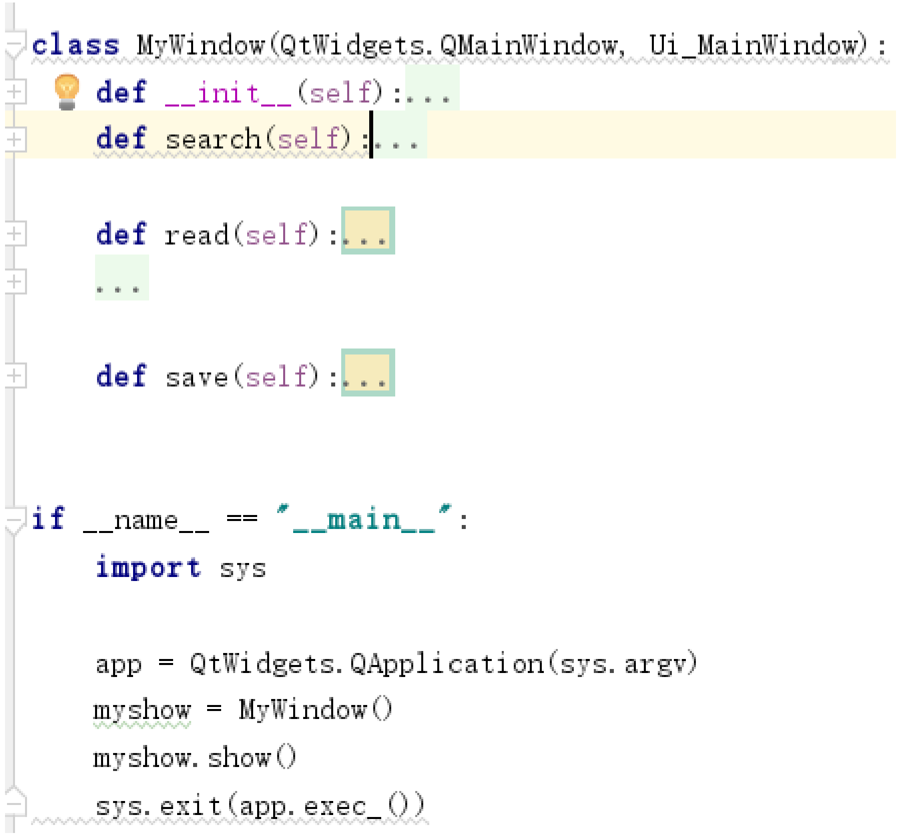
```
pyuic5 -x mainwindow.ui -o mainwindow.py
```

The mainwindow.py including the UI class to initialize the GUI in python.



```python
class Ui_MainWindow(object):
    def setupUi(self, MainWindow):...

    def retranslateUi(self, MainWindow):...
import Images_rc
```

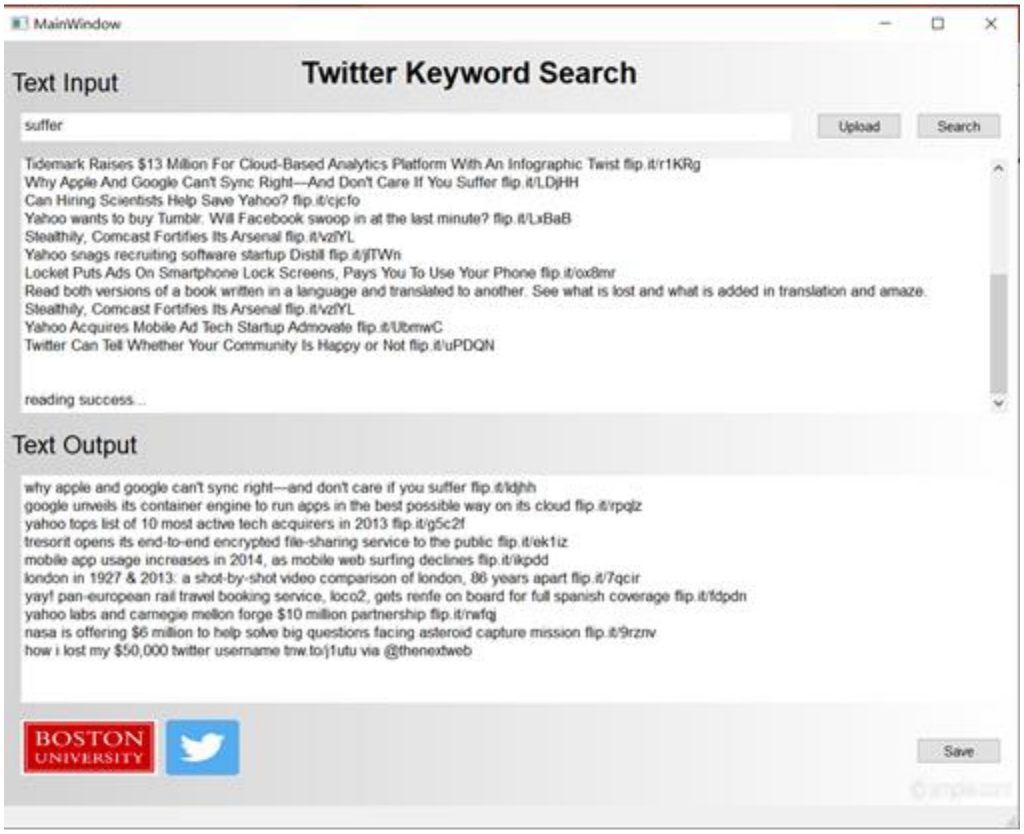
One more step is to combine the keyword Search with the GUI, which means our team needs to build the main script in our project. The main script is as follows:



```python
class MyWindow(QtWidgets.QMainWindow, Ui_MainWindow):
    def __init__(self):...
    def search(self):...

    def read(self):...
    ...

    def save(self):...


if __name__ == "__main__":
    import sys

    app = QtWidgets.QApplication(sys.argv)
    myshow = MyWindow()
    myshow.show()
    sys.exit(app.exec_())
```

The search file is linked with the search button. The read function is linked with the upload button with the read slot. The save is linked with the save button with the save slot.

The final searching result looks like this:



```
Tidemark Raises $13 Million For Cloud-Based Analytics Platform With An Infographic Twist flip.it/r1KRg
Why Apple And Google Can't Sync Right—And Don't Care If You Suffer flip.it/LDjHH
Can Hiring Scientists Help Save Yahoo? flip.it/cjcfo
Yahoo wants to buy Tumblr. Will Facebook swoop in at the last minute? flip.it/LxBaB
Stealthily, Comcast Fortifies Its Arsenal flip.it/vzfYL
Yahoo snags recruiting software startup Distill flip.it/jlTWn
Locket Puts Ads On Smartphone Lock Screens, Pays You To Use Your Phone flip.it/ox8mr
Read both versions of a book written in a language and translated to another. See what is lost and what is added in translation and amaze.
Stealthily, Comcast Fortifies Its Arsenal flip.it/vzfYL
Yahoo Acquires Mobile Ad Tech Startup Admovate flip.it/UbmwC
Twitter Can Tell Whether Your Community Is Happy or Not flip.it/uPDQN


reading success...
```

```
why apple and google can't sync right—and don't care if you suffer flip.it/ldjhh
google unveils its container engine to run apps in the best possible way on its cloud flip.it/rpqlz
yahoo tops list of 10 most active tech acquirers in 2013 flip.it/g5c2f
tresorit opens its end-to-end encrypted file-sharing service to the public flip.it/ek1iz
mobile app usage increases in 2014, as mobile web surfing declines flip.it/ikpdd
london in 1927 & 2013: a shot-by-shot video comparison of london, 86 years apart flip.it/7qcir
yay! pan-european rail travel booking service, loco2, gets renfe on board for full spanish coverage flip.it/fdpdn
yahoo labs and carnegie mellon forge $10 million partnership flip.it/rwfqj
nasa is offering $6 million to help solve big questions facing asteroid capture mission flip.it/9rznv
how i lost my $50,000 twitter username tnw.to/j1utu via @thenextweb
```

Reference:
[1] Wikipedia: https://en.wikipedia.org/wiki/Inverted_index
[2] Qt Designer Manual: https://doc.qt.io/qt-5/qtdesigner-manual.html
[3] River Bank Computing: https://riverbankcomputing.com/software/pyqt/intro