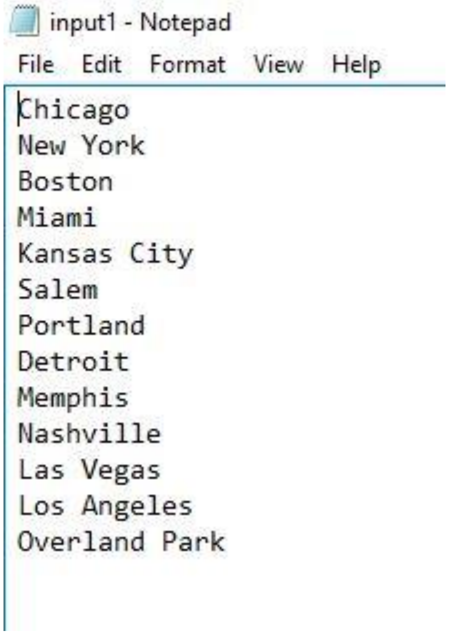
CS5542 Big Data Apps and Analytics

LAB ASSIGNMENT #2

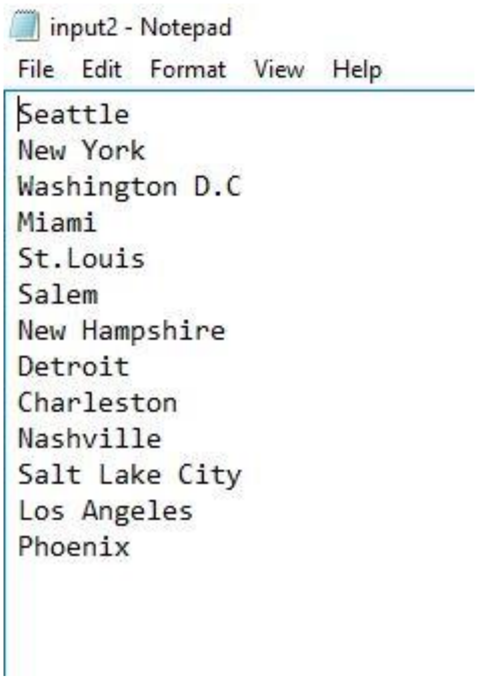
Report

I have used several transformations and actions in the program. These actions and transformations are used to generate output files. These output files are outputs generated from multiple actions and transformations. The object/class **sparkAT** is created and in this the program using several actions and transformations is written. Here, three text files are used as inputs and are combined using **union**.

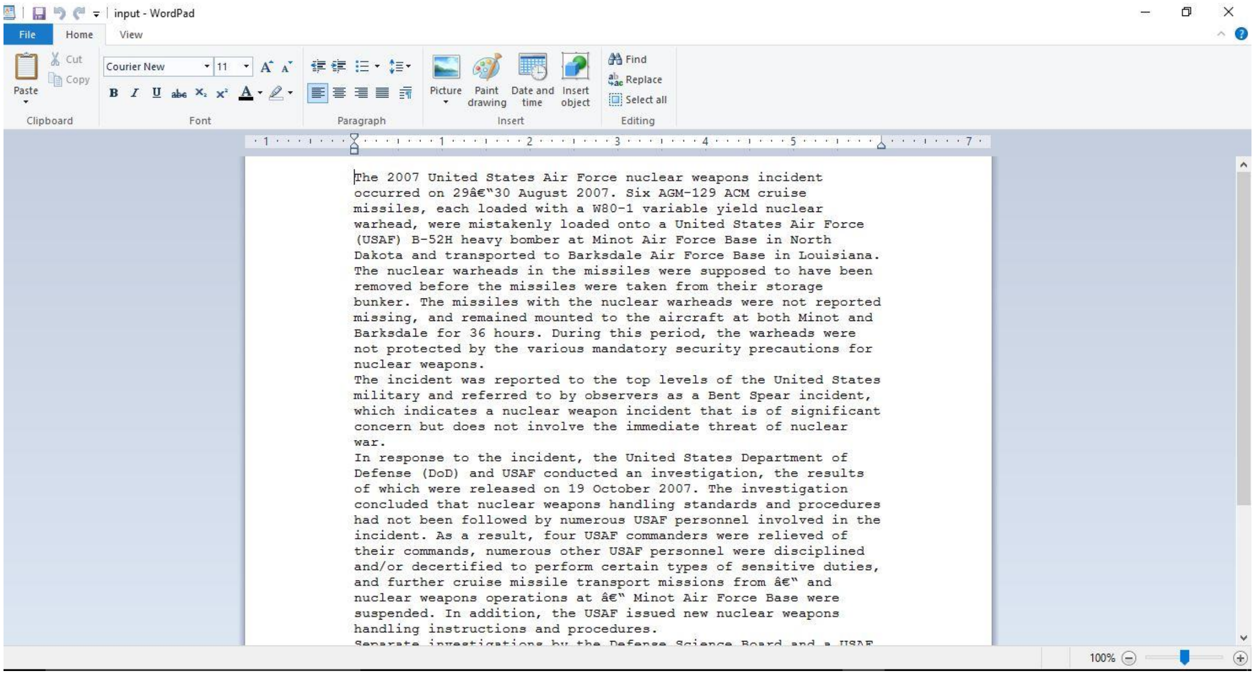
INPUTS:



```
Chicago
New York
Boston
Miami
Kansas City
Salem
Portland
Detroit
Memphis
Nashville
Las Vegas
Los Angeles
Overland Park
```

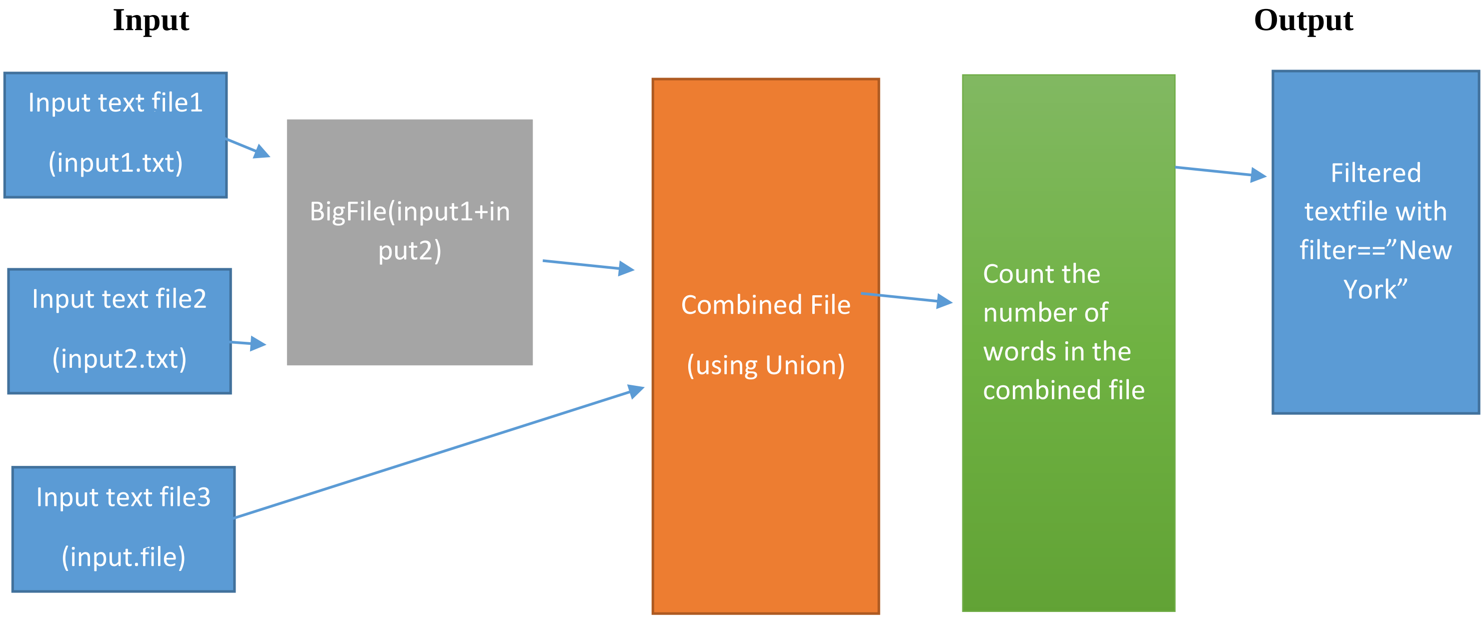

input2 - Notepad
File
Edit
Format
View
Help

```
Seattle
New York
Washington D.C
Miami
St.Louis
Salem
New Hampshire
Detroit
Charleston
Nashville
Salt Lake City
Los Angeles
Phoenix
```


input - WordPad
File
Home
View
Paste
Cut
Copy
Courier New
11
Clipboard
Font
Paragraph
Picture
Paint drawing
Date and time
Insert object
Insert
Find
Replace
Select all
Editing
100%

```
The 2007 United States Air Force nuclear weapons incident
occurred on 29â€“30 August 2007. Six AGM-129 ACM cruise
missiles, each loaded with a W80-1 variable yield nuclear
warhead, were mistakenly loaded onto a United States Air Force
(USAF) B-52H heavy bomber at Minot Air Force Base in North
Dakota and transported to Barksdale Air Force Base in Louisiana.
The nuclear warheads in the missiles were supposed to have been
removed before the missiles were taken from their storage
bunker. The missiles with the nuclear warheads were not reported
missing, and remained mounted to the aircraft at both Minot and
Barksdale for 36 hours. During this period, the warheads were
not protected by the various mandatory security precautions for
nuclear weapons.
The incident was reported to the top levels of the United States
military and referred to by observers as a Bent Spear incident,
which indicates a nuclear weapon incident that is of significant
concern but does not involve the immediate threat of nuclear
war.
In response to the incident, the United States Department of
Defense (DoD) and USAF conducted an investigation, the results
of which were released on 19 October 2007. The investigation
concluded that nuclear weapons handling standards and procedures
had not been followed by numerous USAF personnel involved in the
incident. As a result, four USAF commanders were relieved of
their commands, numerous other USAF personnel were disciplined
and/or decertified to perform certain types of sensitive duties,
and further cruise missile transport missions from â€“ and
nuclear weapons operations at â€“ Minot Air Force Base were
suspended. In addition, the USAF issued new nuclear weapons
handling instructions and procedures.
Separate investigations by the Defense Science Board and a USAF
```

The Diagram of the transformations and actions of the program:


Input
Output
Input text file1
(input1.txt)
Input text file2
(input2.txt)
Input text file3
(input.file)
BigFile(input1+input2)
Combined File
(using Union)
Count the number of words in the combined file
Filtered textfile with filter==”New York”

The output of this particular is saved as text file output.txt, the output files are also provided.

Transformations:

Union()

Distinct()

Intersection()

Filter()

Flatmap()

Actions:

Reduce()

Count()

saveAsTextFile()

collect()