**Input files:**

“invoice.txt”: words of an invoice generated by the OCR engine

“suppliernames.txt”: a list of supplier names

**Run the main program**:

Version 1:

“suppliernames\_finder.py”

(The function is included in the file "bisearch.py")

Version 2:

“suppliernames\_finder2.py” produces the same result, and this program is slightly lengthier. But I think this one has better scalability than the previous version.

In this version, I added some supplier names to the file to make the problem more difficult:

“demo comp”, “demo company2” and “demo company3”.

The new supplier names are in “suppliernames\_mod.txt”.

**Output:**

“word\_list.xlsx”: information extracted from “invoice.txt”

- First column: page number

- Second column: line number

- Third column: position of word

“invoice\_rec.xlsx”: reconstructed invoice from “invoice.txt”

“Sup\_name\_found.xlsx”: a list of company names found in the invoice