# Assignment 1

INFORMATION RETRIVAL

Prabhat Kumar(MT17036)

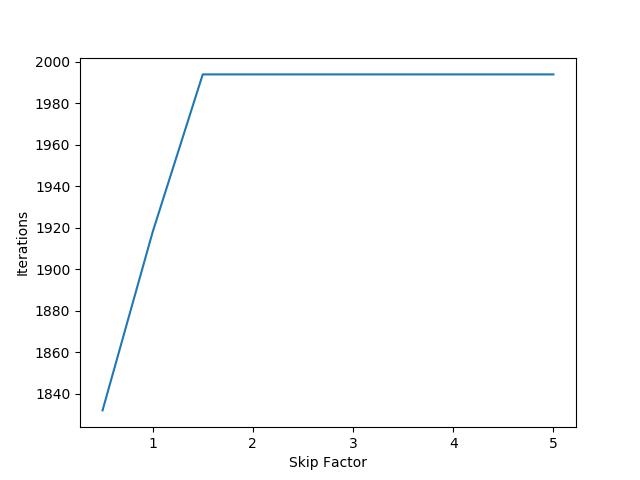
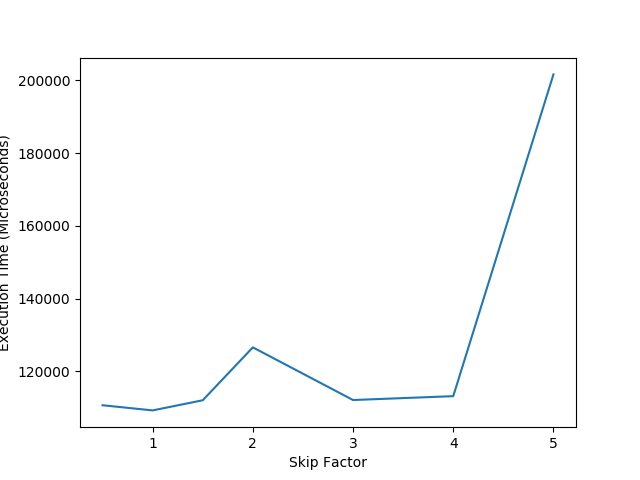
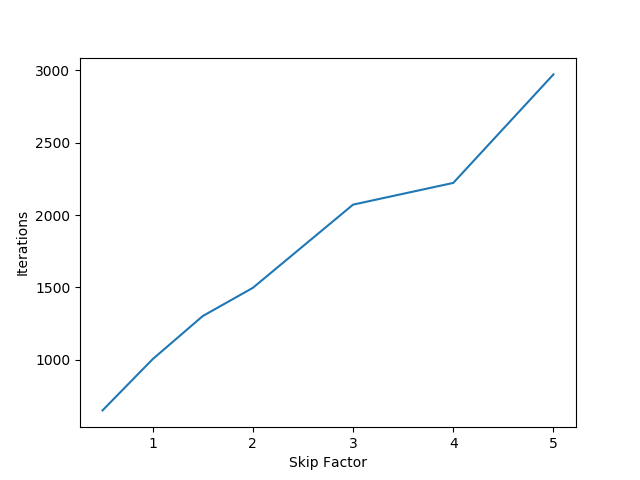
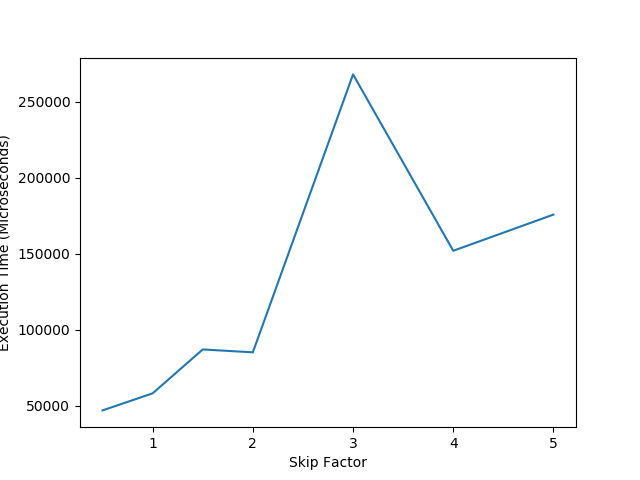
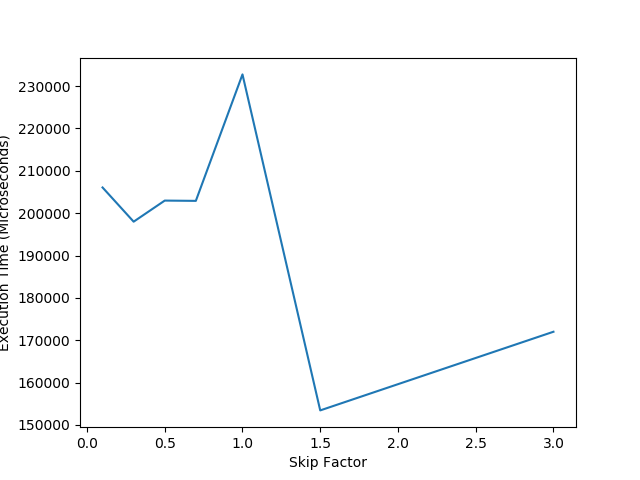
## Analysis Assumptions

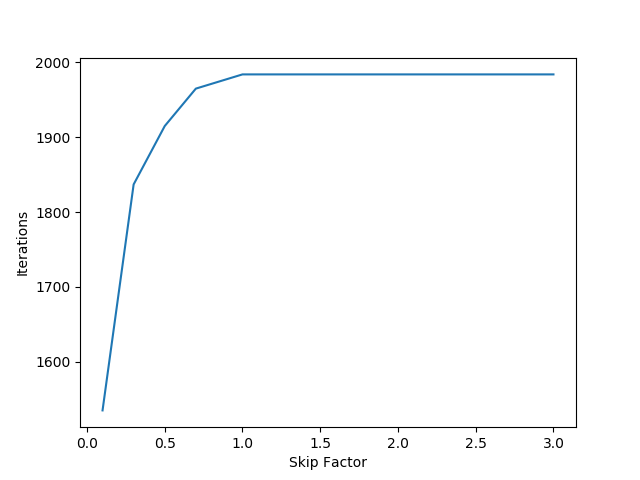
* Comparison has been done on the basis of two parameters – Execution Time and No of Iteration performed against Skip Factor.
* Skip Length has been defined as

Skip Length = Skipfactor \* Sqrt(Length of Posting)

* First Skip Pointer is assumed at Position 0 in the Posting.
* Execution time has been calculated as the difference of System Time returned by “timeit.default\_timer()’’
* Execution Time and No of iteration readings have been taken by averaging the reading over 10 executions.

## Data Observation

1. Great VS Hit
2. Know VS Edu
3. Book VS God



## Inferences

* As skip Factor defines the multiplicative factor of Sqrt(length of posting), it is found that upon increasing the value of Skip Factor , i.e. increasing the skip length tends to make the iteration performed by Skip List Merge tends to those performed by Simple Merge algorithm
* If the occurrence of queried words is high then, the effect of Skip List Merge Algorithm is almost equal to standard Merge algorithm for AND operation.
* Execution Time does not provide concrete evidence of relation of Skip Length. Though Data indicates sharp decrease in Execution time when Skip Length is defined near Sqrt(Length of Posting).
* Execution time also tends to increase as value of Skip Length increases in some cases.

## References

1. WordCloud Repository : <https://amueller.github.io/word_cloud/>
2. NLTK Toolkit : http://www.nltk.org/