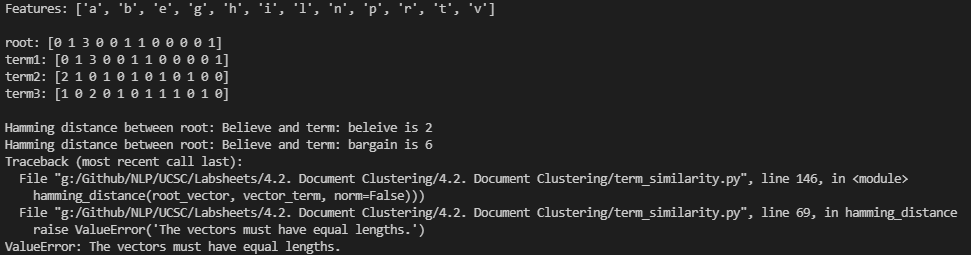
**Lab sheet 9 - Document Clustering**

Kavishka Gamage – 17000475

**Term Similarity**

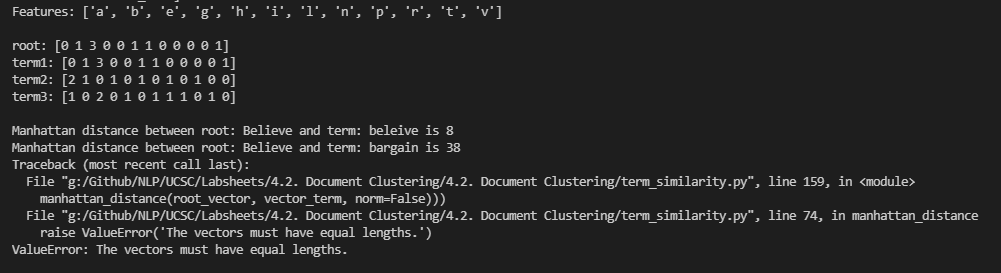
**Hamming distance**

Elephant term gives error when finding text similarity with believe

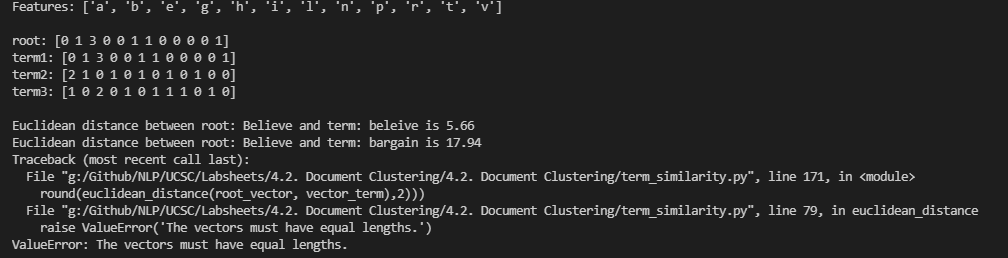
****

**Manhattan distance**

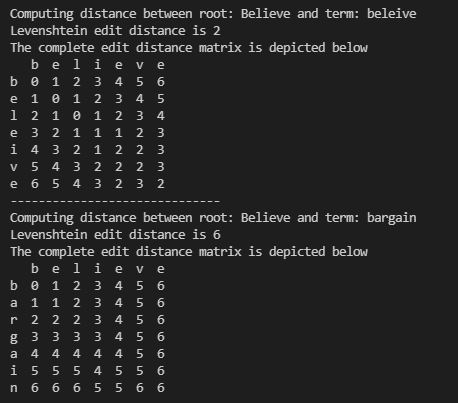
This measure also gave an error when terms don’t have equal length

****

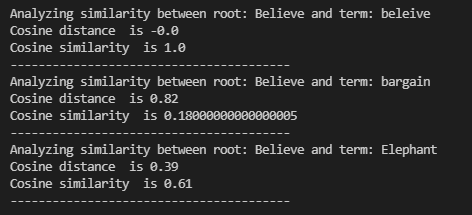
**Euclidian distance**

****

**Levenshtein edit distance**

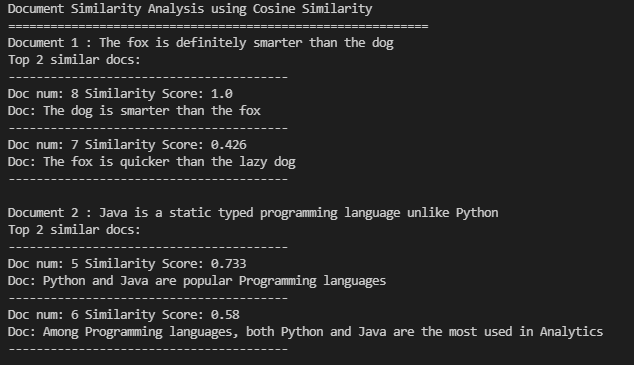


**Cosine similarity**

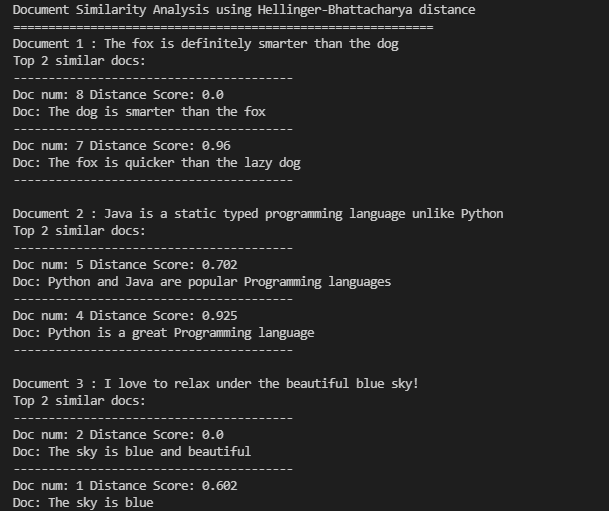
****

**Document Similarity**

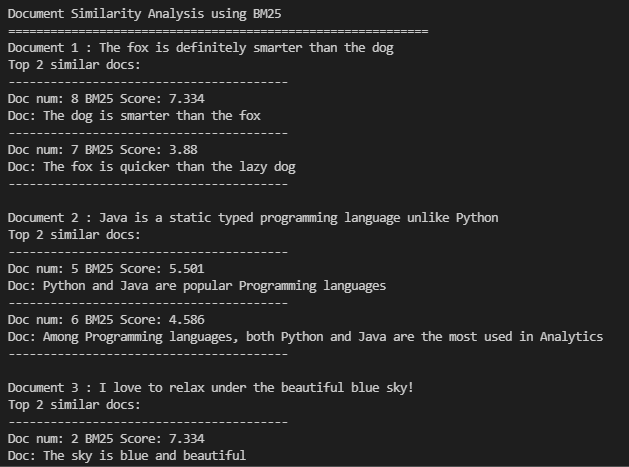
Cosine similarity



**Hellinger-Bhattacharya distance**

****

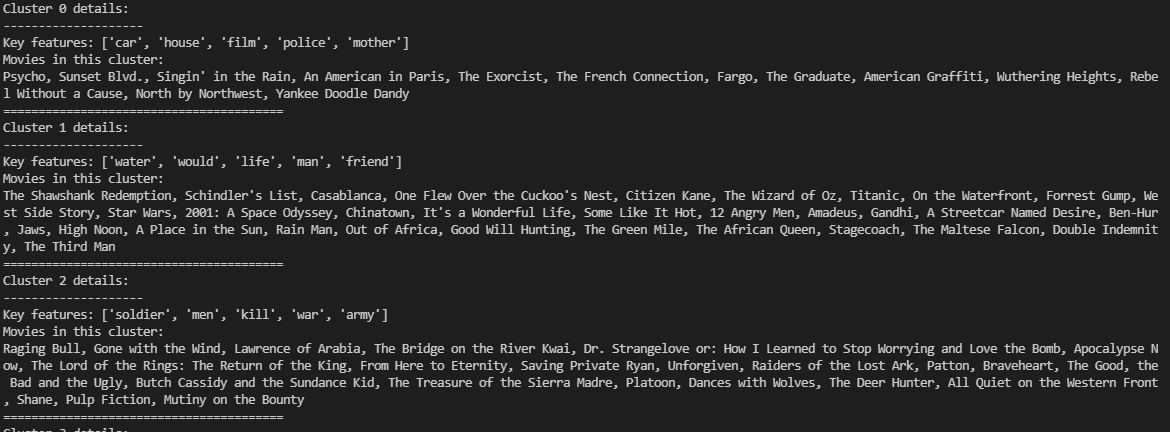
**BM25**

****

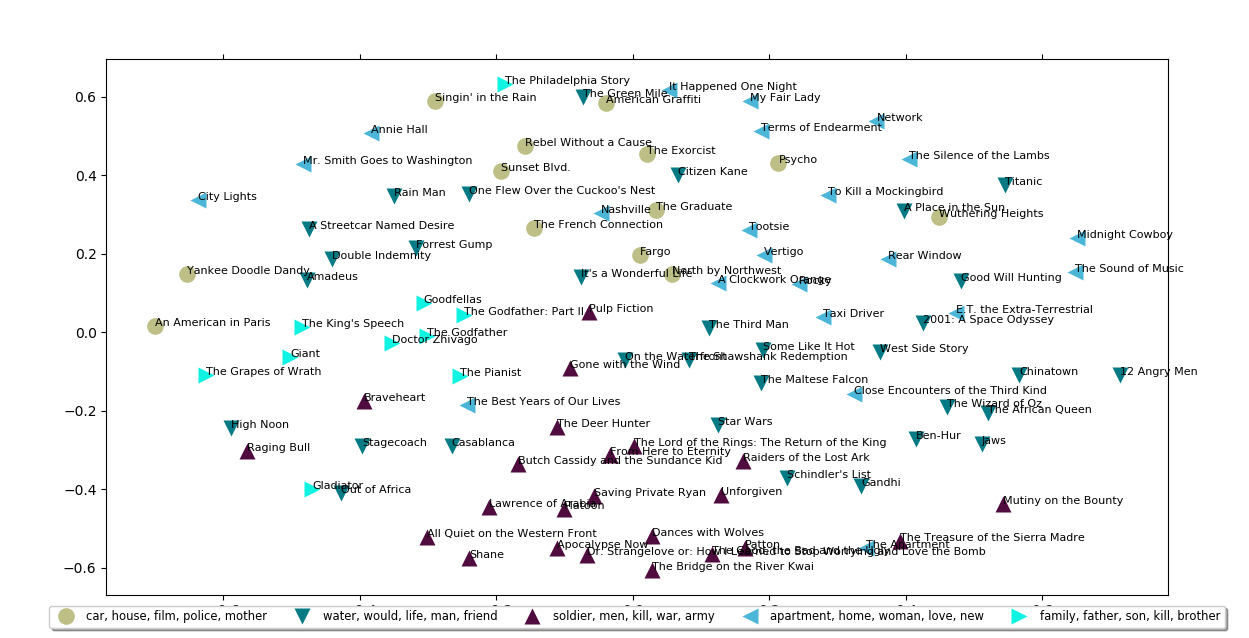
**Document Clustering**

**K-means**

Cluster details

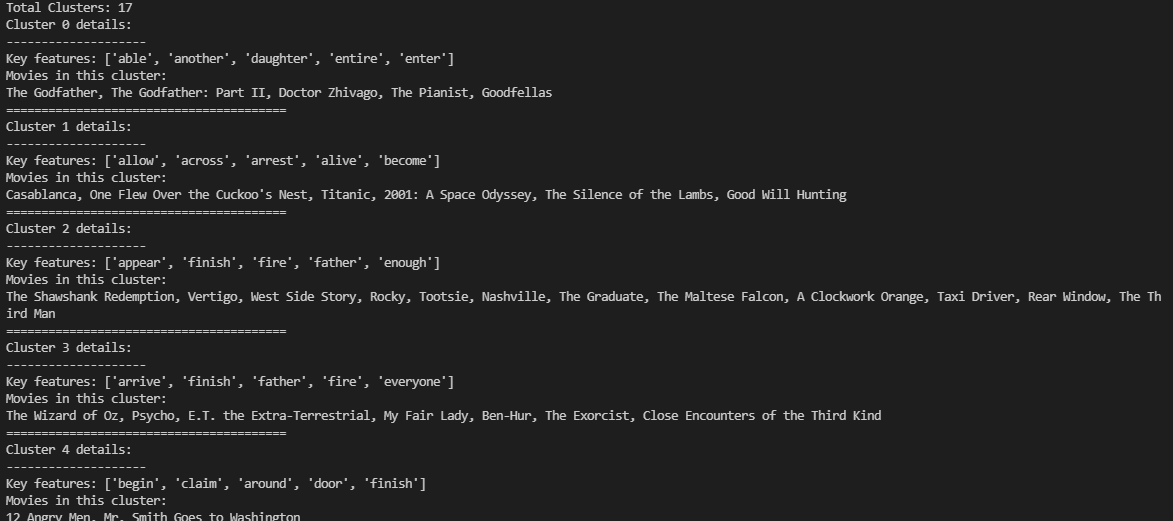


Cluster visualization

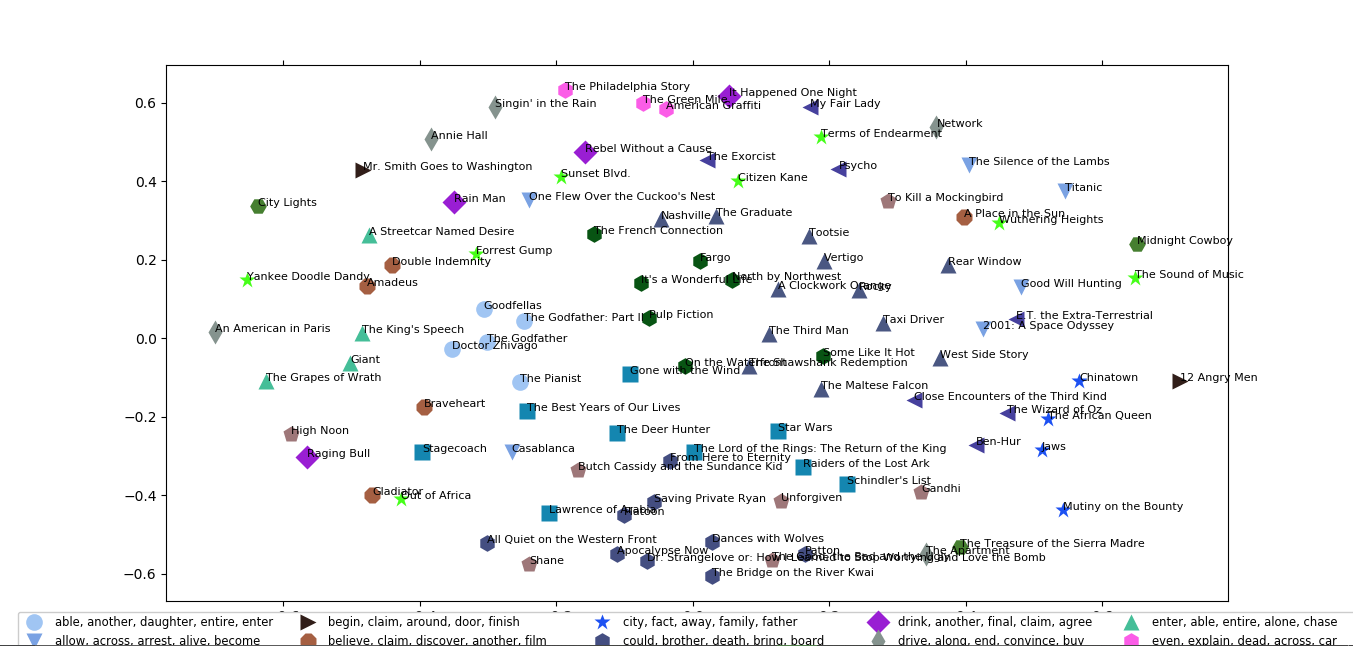
****

**Affinity Propagation**

**Cluster details**

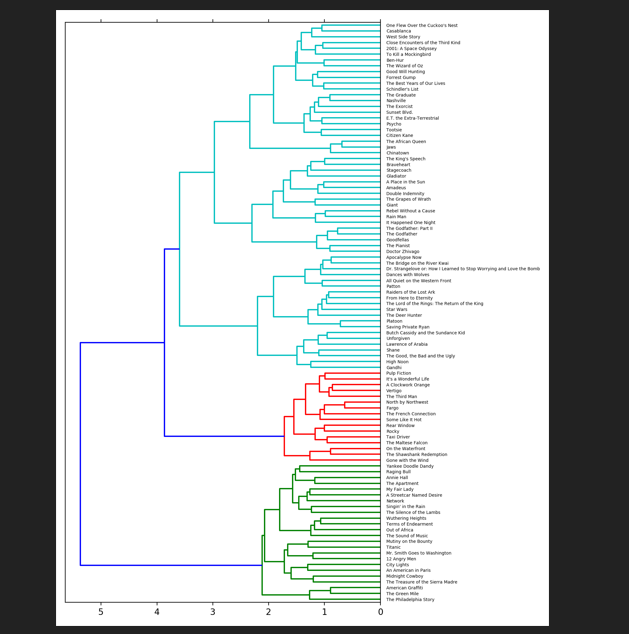
****

**Cluster visualization**

****

**Hierarchical clustering**

**Dendrogram**

****