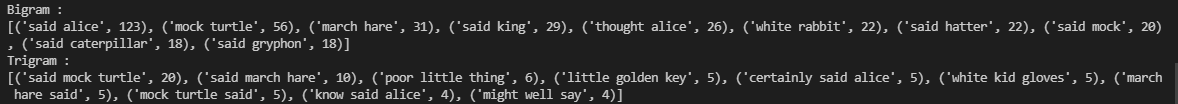
**Lab sheet 11 – Text Summarization**

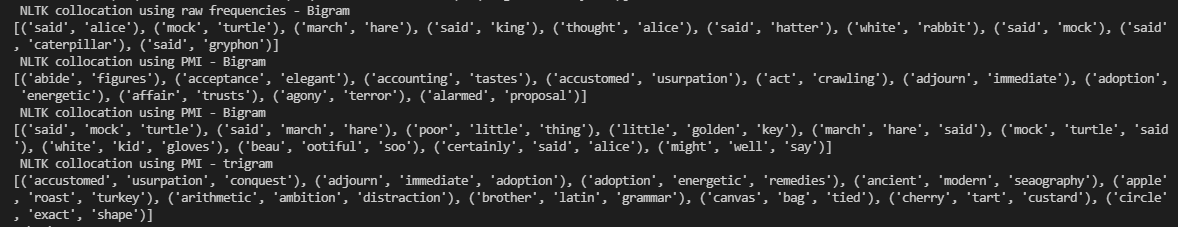
Kavishka Gamage – 17000475

**Key Phrase Extraction**

Collocations N-grams (Bigram and Trigram)

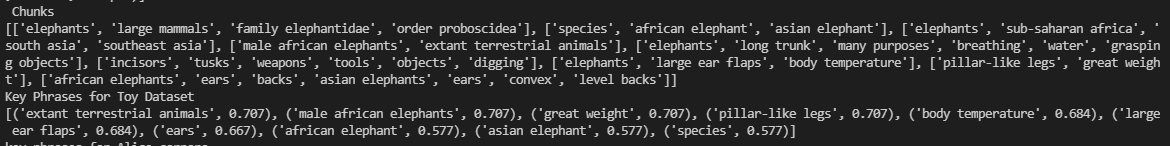


NLTK built in collocation with raw frequency and PMI (Pointwise Mutual Information)

****

Weighted tag based phrase extraction

Toy corpus



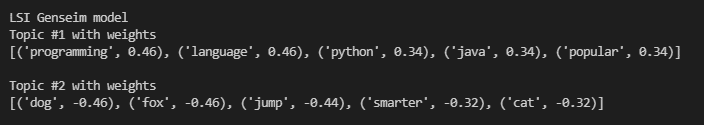
Alice corpus

****

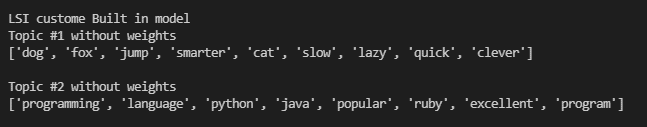
**Topic Modeling**

**For Toy corpus**

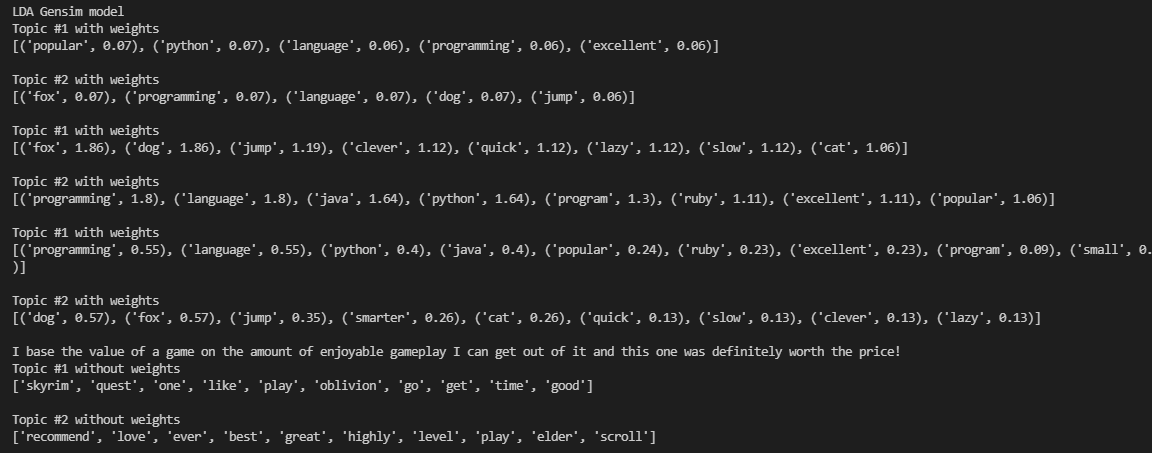
LSI Gensim model - TFIDF



LSI custom built model

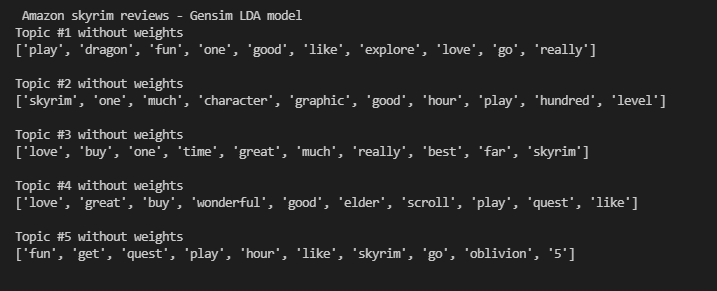


LDA genism model

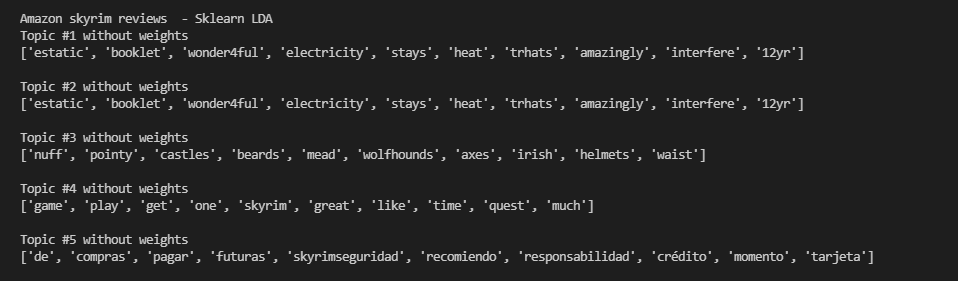


**For Amazon skyrim reviews**

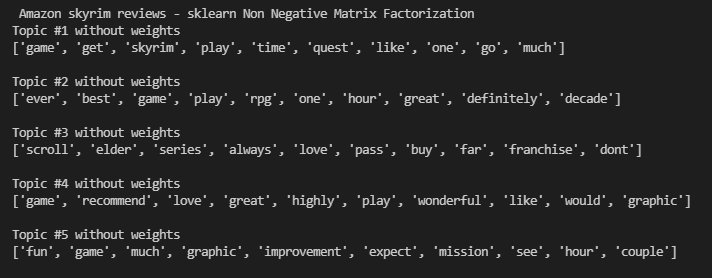
LDA genism model



Sklearn LDA



Sklearn Non Negative Matrix Factorization

****

**Document Summarization**

Gensim summarization model has removed from genism. So its result not included here.

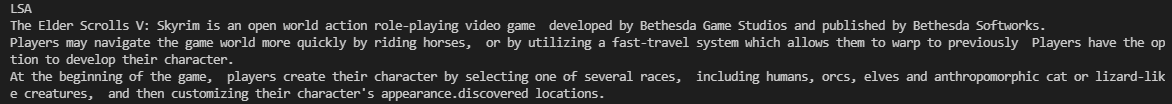
Custom summarization algorithm result



Pagerank - networkX



Generic LSA based text summarizer result



Generic TextRank based text summarizer

