NLP – Cosine Similarity Workshop

Objective

To practice using Cosine Similarity as a tool for performing searching for relevant documents based on a given query string

Exercise

You are given a text file called "quotes.txt" that contains 51 quotes. Given a query string, list out the most relevant quotes.

First, treat each quote as a separate document and compute TFIDF values for each word in the 51 documents (remember to treat each line of quote as a separate document; your corpus is the 51 documents).

Next, treat the query string as a separate document and compute its TFIDF values as well.

Then, perform Cosine Similarity between the query string (treat it as a document) and the 51 documents.

A cosine similarity **greater than 0** is regarded as a **relevant document** for the query string. **Rank the documents** (i.e. lines of quote) by their cosine similarity scores (descending order - higher cosine similarity first)

For pre-processing, remember to **remove stop-words** from your corpus and **stem each word** in the text.

Use the query string – "**life wise choices**" to test your results.

Given the above query string, your results should be:

```
Search results for query: 'life wise choices':

Life is the sum of our choices. [score = 0.5689616110734828]

A wise man will make more opportunities than he finds. [score = 0.303443043025219]

It is our choices that show what we really are, far more than our abilities. [score = 0.27390067640588867]

You attract into your life that which you are. [score = 0.24161974048686424]

A clever person solves a problem; a wise person avoids it. [score = 0.22329624273452417]

Challenges are what make life interesting and overcoming them is what makes life meaningful. [score = 0.19128750621389298]

Let your life be shaped by decisions you made, not by the ones you didn't. [score = 0.11559224862578169]

You can live your whole life and never know who you are; until you see the world through the eyes of others. [score = 0.10677233134042569]

You have enemies? Good. That means you've stood up for something, sometime in your life. [score = 0.10075261917226586]

Our greatest fear should not be of failure, but of succeeding at things in life that don't really matter. [score = 0.09933471565885807]

Don't be afraid your life will end; be afraid that it will never begin. [score = 0.09685798060862479]

There are two things to aim at in life: first, to get what you want; and, after that, to enjoy it. Only the wisest of mankind achieve the second. [score = 0.08538216347679319]

To see the world, things dangerous to come to. To see behind walls, to draw closer. To find each other and to feel. That, is the purpose of life. [score = 0.07666478613699003]
```