NLP - Summarization Workshop

Objective

To practice using TFIDF as a tool for text featurization

Exercise

You are given a text file called "space_invaders.txt". Your task is to produce a 10-line summary of the text.

The algorithm is rather simple. First, compute the TFIDF value of each word in the text file by **treating each sentence as a document** and **the entire text file as the corpus**. Then, sum the **TFIDF values** of **each sentence**. Finally, selects sentences that have the **top 10 TFIDF sums** and print them **in the order that they appear in the text file**.

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

Your summary might look like this:

```
Top TF-IDF sentences: [48, 10, 28, 37, 22, 24, 1, 39, 2, 36]

[Line 1]
The aim is to defeat five rows of eleven aliens—although some versions feature different numbers—that move horizontally
back and forth across the screen as they advance toward the bottom of the screen.

[Line 2]
The player's laser cannon is partially protected by several stationary defense bunkers—the number also varies by
version—that are gradually destroyed from the top and bottom by blasts from either the aliens or the player.

[Line 10]
The game's inspiration is reported to have come from varying sources, including an adaptation of the mechanical game
Space Monsters released by Taito in 1972, and a dream about Japanese school children who are waiting for Santa Claus
when they are attacked by invading aliens.

[Line 22]
Because microcomputers in Japan were not powerful enough at the time to perform the complex tasks involved in designing
and programming Space Invaders, Nishikado had to design his own custom hardware and development tools for the game.

[Line 24]
The game uses an Intel 8080 central processing unit (CPU), displays raster graphics on a CRT monitor using a bitmapped
framebuffer, and uses monaural sound hosted by a combination of analog circuitry and a Texas Instruments SN76477 sound
chip.

[Line 28]
Despite the specially developed hardware, Nishikado was unable to program the game as he wanted—the Control Program
board was not powerful enough to display the graphics in color or move the enemies faster—and he ended up considering
the development of the game's hardware the most difficult part of the whole process.

[Line 36]
The cabinet artwork featured large humanoid monsters not present in the game; Nishikado attributes this to the artist
basing the designs on the original title of "Space Monsters", rather than referring to the actual in-game graphics.

[Line 37]
In the upright cabinets, the game graphics are generated on a hidden CRT monitor and reflected toward the player using
a semi-transparent mirror, behind which is mounted a plastic cutout of a moon bolted against a painted starry
background.

[Line 39]
Both Taito's and Midway's first Space Invaders versions had black-and-white graphics with a transparent colored overlay
using strips of orange and green cellophane over certain portions of the screen to add color to the image.

[Line 48]
It was also the first game where players were given multiple lives, had to repel hordes of enemies, could take cover
from enemy fire, and use destructible barriers,in addition to being the first game to use a continuous background
soundtrack, with four simple diatonic descending bass notes repeating in a loop, which was dynamic and changed pace
during stages, like a heartbeat sound that increases pace as enemies approached.
```