Implement the TF-IDF Algorithm from Scratch

Your task is to implement the Term Frequency-Inverse Document Frequency (TF-IDF) algorithm from scratch in Python, without using any pre-built packages or libraries. You may use NLTK or other libraries for tokenization, stemming, and other preprocessing tasks.

Background

TF-IDF is a numerical statistic that is used to reflect the importance of a term in a document in a collection or corpus of documents. The TF-IDF value increases proportionally to the number of times a word appears in the document, but is offset by the frequency of the word in the corpus, which helps to adjust for the fact that some words appear more frequently in general.

The TF-IDF algorithm has two parts:

1. **Term Frequency (TF)**: The number of times a term appears in a document, divided by the total number of terms in the document. This value is often normalized to prevent bias towards longer documents.
2. **Inverse Document Frequency (IDF)**: The logarithmically scaled inverse fraction of the documents that contain the word, obtained by dividing the total number of documents by the number of documents containing the term, and then taking the logarithm of that quotient.

The TF-IDF value for a term in a document is then calculated as:

*TF-IDF(term, document, corpus) = TF(term, document) * IDF(term, corpus)*

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

$tf_{i,j}$ = number of occurrences of i in j

df_i = number of documents containing i

N = total number of documents

Requirements

Your implementation should include the following:

1. A function to tokenize a document into a list of words.
2. A function to compute the term frequency of each term in a document.
3. A function to compute the inverse document frequency of each term in the corpus.
4. A function to compute the TF-IDF score for each term in each document in the corpus.

You may assume that the corpus is a list of documents.

You should use this corpus:

corpus = [

"Once upon a time, in a faraway land, there was a brave knight named Sir Lancelot. He was known for his strength, courage, and chivalry. One day, the King of the land asked Sir Lancelot to rescue his daughter, who had been kidnapped by an evil sorcerer. Sir Lancelot set out on his quest, facing many dangers along the way. But he never gave up, and eventually he rescued the princess and defeated the sorcerer. The people of the land cheered for Sir Lancelot, and he became a legend in his own time.",

"In a village at the foot of a mountain, there lived a poor farmer named Jack. He had a small farm and a cow, which was his only possession. One day, the cow stopped giving milk, and Jack didn't know what to do. So he decided to sell the cow at the market. On the way, he met a stranger who offered to trade five magic beans for the cow. Jack agreed, and when he got home, his mother was furious. But that night, the magic beans grew into a giant beanstalk, and Jack climbed it to find a castle in the clouds. There, he met a giant who had a goose that laid golden eggs. Jack stole the goose and ran down the beanstalk, but the giant followed him. Jack chopped down the beanstalk, and the giant fell to his death. Jack and his mother lived happily ever after with the golden eggs.",

"In a kingdom ruled by a wicked queen, there lived a beautiful princess named Snow White. The queen was jealous of Snow White's beauty, and ordered a huntsman to kill her. But the huntsman couldn't do it, so he left Snow White in the forest. There, she met seven dwarfs who took her in and cared for her. But the queen found out that Snow White was still alive, and disguised herself as an old woman to give Snow White a poisoned apple. Snow White fell into a deep sleep, but a prince came and woke her with a kiss. They lived happily ever after, and the queen got what she deserved.",

"In a world of magic and wonder, there was a young wizard named Harry Potter. He had been orphaned as a baby, and was raised by his cruel relatives. But one day, he received a letter from Hogwarts School of Witchcraft and Wizardry, inviting him to attend. There, he learned about his true heritage and his destiny to defeat the dark wizard Voldemort. Harry made many friends at Hogwarts, including Hermione Granger and Ron Weasley. Together, they faced many challenges and battles, but in the end, Harry was able to vanquish Voldemort and bring peace to the wizarding world.",

"In a land of dragons and knights, there was a beautiful princess named Fiona. She had been cursed by a wicked sorcerer and turned into an ogre. One day, a brave knight named Shrek was sent to rescue her from a tower. But when he found her, he discovered that she was an ogre. They didn't get along at first, but eventually they fell in love. Along the way, they met many fairy tale characters, including a talking donkey and a gingerbread man. Together, they defeated the evil Lord Farquaad and lived happily ever after in the swamp."

]

Evaluation

Your implementation will be evaluated based on the correctness of the TF-IDF scores it produces, as well as the efficiency and readability of your code.

You will be asked to explain your code and decisions as well as any conclusions/outputs during your next interview.

Additional Notes

Python programming language is preferred. You may not use any pre-built packages or libraries for the TF-IDF algorithm, but you may use libraries for other tasks such as tokenization and data manipulation.