SUYASH PRATAP SINGH(181B226) ¶

TASKS:-

1. Implement the Cosine similarity based Vector method of Document ranking.
2. For reference take the example from the lecture slides and check the output to be same.
3. The output should be the order of documents and their ranks value.

```python
In [1]: from sklearn.feature_extraction.text import TfidfVectorizer
        from sklearn.metrics.pairwise import linear_kernel

        search_terms = 'gold silver truck'
        documents = ['Shipment of gold damaged in a fire', 'Delivery of silver arrived
        in a silver truck','Shipment of gold arrived in a truck']

        doc_vectors = TfidfVectorizer().fit_transform([search_terms] + documents)

        cosine_similarities = linear_kernel(doc_vectors[0:1], doc_vectors).flatten()
        document_scores = [item.item() for item in cosine_similarities[1:]]
        print(cosine_similarities)
        print(document_scores)
```

```
[1.         0.17338464 0.59652612 0.40111381]
[0.17338463517555858, 0.5965261170486699, 0.4011138131157242]
```

Conclusion:-

From the Document_scores and cosine_similarities the rank is D2>D3>D1

THANK YOU