Expno: 5 NLP Tat: part as spa tagging and information

Aim:-

TO perform part of speech tagging on a given text using spacy and develop an information retrieval system that ranks a set of document based the relevance to a user query using tf-idf vectorization and cosine similarities

Algorithm:-

1. Load the spacy English model and input the text of POS tagging
2. Process the text to tokens and assign part-of speech to each token
3. Collect document and a query the combine them into a single vector
4. Vectorize the corpus using TF-IDF and compute cosine similarity between the query and documents

Program:

```
import pandas as pd
import re.
import spacy
nlp = spacy.load("en_core_web_sm")
text = " "
doc = nlp (text)
for token in doc:
    print (f"{token: token:15} ->"3
from sklearn, feature_extraction
from sklearn. metrics
```

Output:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 8807 entries, 0 to 8806

data columns (total 12 columns)

#	Column	Non-Null Count	Dtype
0	type	8807	non-null
1	title	8807	non-null
2	director	8173	non-null
3	cast	7982	non-null
4	country	7976	non-null
5	date_added	8797	non-null
6	release_year	8807	non-null
7	rating	8803	non-null
8	duration	8804	non-null

documents = ["AI tools analyses student performance and provide feedback"
"Intelligent tutoring system adapt to each student learning style"]

similarities = cosine_similarity(tfidf_matrix(q))
ranked_docs = sorted(zip(sc, docs), reverse=True)
for score, doc in ranked_res:
print("{score}: {score} -> {doc}")

Result:

This program has been executed successfully