xkqmpzzui

October 9, 2024

[1]:
```python
# Define the documents
document1 = "The quick brown fox jumped over the lazy dog."
document2 = "The lazy dog slept in the sun."

# Step 1: Tokenize the documents
# Convert each document to lowercase and split it into words
tokens1 = document1.lower().split()
tokens2 = document2.lower().split()

# Combine the tokens into a list of unique terms
terms = list(set(tokens1 + tokens2))

# Step 2: Build the inverted index
# Create an empty dictionary to store the inverted index
inverted_index = {}

# For each term, find the documents that contain it
for term in terms:
    documents = []
    if term in tokens1:
        documents.append("Document 1")
    if term in tokens2:
        documents.append("Document 2")
    inverted_index[term] = documents

# Step 3: Print the inverted index
for term, documents in inverted_index.items():
    print(term, "->", ", ".join(documents))
```

```
lazy -> Document 1, Document 2
sun. -> Document 2
dog -> Document 2
dog. -> Document 1
brown -> Document 1
slept -> Document 2
in -> Document 2
fox -> Document 1
over -> Document 1
```

```
jumped -> Document 1
quick -> Document 1
the -> Document 1, Document 2
```

[ ]: