```
D1 = "The weather is sunny today"
D2 = "I love this new phone"
D3 = "The meeting starts at noon"
D4 = "She feels tired and stressed"
D5 = "The package arrived late"
D6 = "He won the tennis match"
D7 = "This restaurant is amazing"
D8 = "The service was terrible"
D9 = "I am learning natural language processing"
D10 = "The movie was boring"
D11 = "Prices increased last month"
D12 = "She adopted a cute puppy"
D13 = "The system crashed unexpectedly"
D14 = "We traveled to Paris"
D15 = "The book is very interesting"
D16 = "He forgot his password"
D17 = "The coffee tastes great"
D18 = "Traffic was heavy this morning"
D19 = "She passed the exam"
D20 = "The room is clean and quiet"
```

```
from sklearn.feature_extraction.text import CountVectorizer
import pandas as pd

documents = [D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18, D19, D20]

# Initialize TfidfVectorizer
vectorizer = CountVectorizer(max_df=0.95, min_df=1, stop_words="english")

# Fit and transform the documents
bow = vectorizer.fit_transform(documents)

# Get feature names (words)
feature_names = vectorizer.get_feature_names_out()

# Convert to DataFrame for better readability
df_bow = pd.DataFrame(bow.toarray(), columns=feature_names)

print("Bag of Words (TF-IDF) Representation:")
print(df_bow)
```

```
Bag of Words (TF-IDF) Representation:
    adopted  amazing  arrived  book  boring  clean  coffee  crashed  cute  \
0         0        0        0     0       0      0       0        0     0
1         0        0        0     0       0      0       0        0     0
2         0        0        0     0       0      0       0        0     0
3         0        0        0     0       0      0       0        0     0
4         0        0        1     0       0      0       0        0     0
5         0        0        0     0       0      0       0        0     0
6         0        1        0     0       0      0       0        0     0
7         0        0        0     0       0      0       0        0     0
8         0        0        0     0       0      0       0        0     0
9         0        0        0     0       1      0       0        0     0
10        0        0        0     0       0      0       0        0     0
11        1        0        0     0       0      0       0        0     1
12        0        0        0     0       0      0       0        1     0
13        0        0        0     0       0      0       0        0     0
14        0        0        0     1       0      0       0        0     0
15        0        0        0     0       0      0       0        0     0
16        0        0        0     0       0      0       1        0     0
17        0        0        0     0       0      0       0        0     0
18        0        0        0     0       0      0       0        0     0
19        0        0        0     0       0      1       0        0     0

    exam  ...  tastes  tennis  terrible  tired  today  traffic  traveled  \
0      0  ...       0       0         0      0      1        0         0
1      0  ...       0       0         0      0      0        0         0
2      0  ...       0       0         0      0      0        0         0
3      0  ...       0       0         0      1      0        0         0
4      0  ...       0       0         0      0      0        0         0
5      0  ...       0       1         0      0      0        0         0
6      0  ...       0       0         0      0      0        0         0
7      0  ...       0       0         1      0      0        0         0
8      0  ...       0       0         0      0      0        0         0
9      0  ...       0       0         0      0      0        0         0
10     0  ...       0       0         0      0      0        0         0
11     0  ...       0       0         0      0      0        0         0
```

```
12     0  ...       0       0         0      0      0        0         0
13     0  ...       0       0         0      0      0        0         1
14     0  ...       0       0         0      0      0        0         0
15     0  ...       0       0         0      0      0        0         0
16     0  ...       1       0         0      0      0        0         0
17     0  ...       0       0         0      0      0        1         0
18     1  ...       0       0         0      0      0        0         0
19     0  ...       0       0         0      0      0        0         0

    unexpectedly  weather  won
0              0        1    0
1              0        0    0
2              0        0    0
3              0        0    0
4              0        0    0
5              0        0    1
6              0        0    0
7              0        0    0
8              0        0    0
9              0        0    0
10             0        0    0
```

```
from sklearn.metrics.pairwise import cosine_similarity
import pandas as pd

# Calculate cosine similarity between documents
cosine_sim_matrix = cosine_similarity(df_bow)

# Convert to DataFrame for better readability
df_cosine_sim = pd.DataFrame(cosine_sim_matrix, index=[D1,D2,D3,D4,D5,D6,D7,D8,D9,D10,D11,D12,D13,D14,D15,D16,D17,D18,D19,D20],

print("Cosine Similarity Matrix:")
print(df_cosine_sim)
```

```
Cosine Similarity Matrix:
                                        The weather is sunny today  \
The weather is sunny today                                     1.0
I love this new phone                                          0.0
The meeting starts at noon                                     0.0
She feels tired and stressed                                   0.0
The package arrived late                                       0.0
He won the tennis match                                        0.0
This restaurant is amazing                                     0.0
The service was terrible                                       0.0
I am learning natural language processing                      0.0
The movie was boring                                           0.0
Prices increased last month                                    0.0
She adopted a cute puppy                                       0.0
The system crashed unexpectedly                                0.0
We traveled to Paris                                           0.0
The book is very interesting                                   0.0
He forgot his password                                         0.0
The coffee tastes great                                        0.0
Traffic was heavy this morning                                 0.0
She passed the exam                                            0.0
The room is clean and quiet                                    0.0

                                        I love this new phone  \
The weather is sunny today                                0.0
I love this new phone                                     1.0
The meeting starts at noon                                0.0
She feels tired and stressed                              0.0
The package arrived late                                  0.0
He won the tennis match                                   0.0
This restaurant is amazing                                0.0
The service was terrible                                  0.0
I am learning natural language processing                 0.0
The movie was boring                                      0.0
Prices increased last month                               0.0
She adopted a cute puppy                                  0.0
The system crashed unexpectedly                           0.0
We traveled to Paris                                      0.0
The book is very interesting                              0.0
He forgot his password                                    0.0
The coffee tastes great                                   0.0
Traffic was heavy this morning                            0.0
She passed the exam                                       0.0
The room is clean and quiet                               0.0

                                        The meeting starts at noon  \
The weather is sunny today                                     0.0
I love this new phone                                          0.0
The meeting starts at noon                                     1.0
```

```
She feels tired and stressed                                0.0
The package arrived late                                    0.0
He won the tennis match                                     0.0
This restaurant is amazing                                  0.0
The service was terrible                                    0.0
I am learning natural language processing                   0.0
The movie was boring                                        0.0
Prices increased last month                                 0.0
She adopted a cute puppy                                    0.0
```

```
from sklearn.metrics import jaccard_score
import numpy as np
import pandas as pd

# Convert BOW matrix to binary (presence/absence) for Jaccard similarity
binary_bow = (df_bow > 0).astype(int)

# Calculate Jaccard Similarity
num_documents = binary_bow.shape[0]
jaccard_sim_matrix = np.zeros((num_documents, num_documents))

for i in range(num_documents):
    for j in range(num_documents):
        # For Jaccard similarity, we can use the jaccard_score which computes 1 - Jaccard distance
        # However, jaccard_score expects 1D arrays for binary classification tasks.
        # To compute similarity between two sets represented as binary vectors,
        # we can calculate (intersection_size / union_size)

        # Get the binary vectors for documents i and j
        vec_i = binary_bow.iloc[i]
        vec_j = binary_bow.iloc[j]

        # Calculate intersection and union
        intersection = np.sum(np.logical_and(vec_i, vec_j))
        union = np.sum(np.logical_or(vec_i, vec_j))

        if union == 0:
            jaccard_sim_matrix[i, j] = 0.0 # If both sets are empty, similarity is 0
        else:
            jaccard_sim_matrix[i, j] = intersection / union

# Convert to DataFrame for better readability
df_jaccard_sim = pd.DataFrame(jaccard_sim_matrix, index=documents, columns=documents)

print("Jaccard Similarity Matrix:")
print(df_jaccard_sim)
```

```
Jaccard Similarity Matrix:
                                        The weather is sunny today  \
The weather is sunny today                                     1.0
I love this new phone                                          0.0
The meeting starts at noon                                     0.0
She feels tired and stressed                                   0.0
The package arrived late                                       0.0
He won the tennis match                                        0.0
This restaurant is amazing                                     0.0
The service was terrible                                       0.0
I am learning natural language processing                      0.0
The movie was boring                                           0.0
Prices increased last month                                    0.0
She adopted a cute puppy                                       0.0
The system crashed unexpectedly                                0.0
We traveled to Paris                                           0.0
The book is very interesting                                   0.0
He forgot his password                                         0.0
The coffee tastes great                                        0.0
Traffic was heavy this morning                                 0.0
She passed the exam                                            0.0
The room is clean and quiet                                    0.0

                                        I love this new phone  \
The weather is sunny today                                0.0
I love this new phone                                     1.0
The meeting starts at noon                                0.0
She feels tired and stressed                              0.0
The package arrived late                                  0.0
He won the tennis match                                   0.0
This restaurant is amazing                                0.0
The service was terrible                                  0.0
I am learning natural language processing                 0.0
The movie was boring                                      0.0
```

```
Prices increased last month                              0.0
She adopted a cute puppy                                 0.0
The system crashed unexpectedly                          0.0
We traveled to Paris                                     0.0
The book is very interesting                             0.0
He forgot his password                                   0.0
The coffee tastes great                                  0.0
Traffic was heavy this morning                           0.0
She passed the exam                                      0.0
The room is clean and quiet                              0.0

                                         The meeting starts at noon  \
The weather is sunny today                                       0.0
I love this new phone                                            0.0
The meeting starts at noon                                       1.0
She feels tired and stressed                                     0.0
The package arrived late                                         0.0
He won the tennis match                                          0.0
This restaurant is amazing                                       0.0
The service was terrible                                         0.0
I am learning natural language processing                        0.0
The movie was boring                                             0.0
Prices increased last month                                      0.0
She adopted a cute puppy                                         0.0
```

```
D1 = "The weather is sunny today"
D2 = "I love this new phone"
D3 = "The meeting starts at noon"
D4 = "She feels tired and stressed"
D5 = "The package arrived late"
D6 = "He won the tennis match"
D7 = "This restaurant is amazing"
D8 = "The service was terrible"
D9 = "I am learning natural language processing"
D10 = "The movie was boring"
D11 = "Prices increased last month"
D12 = "She adopted a cute puppy"
D13 = "The system crashed unexpectedly"
D14 = "We traveled to Paris"
D15 = "The book is very interesting"
D16 = "He forgot his password"
D17 = "The coffee tastes great"
D18 = "Traffic was heavy this morning"
D19 = "She passed the exam"
D20 = "The room is clean and quiet"
import nltk
from nltk.corpus import wordnet
from nltk.tokenize import word_tokenize
import numpy as np
import pandas as pd

try:
    nltk.data.find('corpora/wordnet')
except LookupError:
    nltk.download('wordnet')
try:
    nltk.data.find('tokenizers/punkt_tab')
except LookupError:
    nltk.download('punkt_tab')

def document_wordnet_similarity(doc1, doc2):
    tokens1 = word_tokenize(doc1.lower())
    tokens2 = word_tokenize(doc2.lower())

    synsets1 = [s for token in tokens1 for s in wordnet.synsets(token)]
    synsets2 = [s for token in tokens2 for s in wordnet.synsets(token)]

    if not synsets1 or not synsets2:
        return 0.0

     # Calculate pairwise similarity and take the maximum for each pair
    max_similarities = []
    for s1 in synsets1:
        max_sim_for_s1 = 0.0
        for s2 in synsets2:
            sim = s1.path_similarity(s2)
            if sim is not None and sim > max_sim_for_s1:
                max_sim_for_s1 = sim
        max_similarities.append(max_sim_for_s1)
```

```
    # Average the maximum similarities (simple approach)
    if max_similarities:
        return np.mean(max_similarities)
    else:
        return 0.0

# Prepare documents list
documents = [D1, D2, D3, D4, D5, D6, D7, D8, D9, D10, D11, D12, D13, D14, D15, D16, D17, D18, D19, D20]

# Calculate WordNet similarity matrix
num_documents = len(documents)
wordnet_sim_matrix = np.zeros((num_documents, num_documents))

for i in range(num_documents):
    for j in range(num_documents):
        if i == j:
            wordnet_sim_matrix[i, j] = 1.0
        else:
            wordnet_sim_matrix[i, j] = document_wordnet_similarity(documents[i], documents[j])

df_wordnet_sim = pd.DataFrame(wordnet_sim_matrix, index=documents, columns=documents)

print("WordNet Similarity Matrix (Path Similarity):")
print(df_wordnet_sim)
```

```
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
WordNet Similarity Matrix (Path Similarity):
                                        The weather is sunny today  \
The weather is sunny today                                1.000000
I love this new phone                                     0.226945
The meeting starts at noon                                0.216242
She feels tired and stressed                              0.240832
The package arrived late                                  0.278571
He won the tennis match                                   0.187278
This restaurant is amazing                                0.794444
The service was terrible                                  0.456407
I am learning natural language processing                 0.392605
The movie was boring                                      0.709780
Prices increased last month                               0.220407
She adopted a cute puppy                                  0.172204
The system crashed unexpectedly                           0.185550
We traveled to Paris                                      0.223077
The book is very interesting                              0.479420
He forgot his password                                    0.213384
The coffee tastes great                                   0.192063
Traffic was heavy this morning                            0.443946
She passed the exam                                       0.240217
The room is clean and quiet                               0.426510

                                        I love this new phone  \
The weather is sunny today                           0.251835
I love this new phone                                1.000000
The meeting starts at noon                           0.214880
She feels tired and stressed                         0.226261
The package arrived late                             0.285854
He won the tennis match                              0.216065
This restaurant is amazing                           0.274537
The service was terrible                             0.211862
I am learning natural language processing            0.296241
The movie was boring                                 0.247975
Prices increased last month                          0.215329
She adopted a cute puppy                             0.197076
The system crashed unexpectedly                      0.195007
We traveled to Paris                                 0.226549
The book is very interesting                         0.227502
He forgot his password                               0.277778
The coffee tastes great                              0.203704
Traffic was heavy this morning                       0.274788
She passed the exam                                  0.230601
The room is clean and quiet                          0.269943

                                        The meeting starts at noon  \
The weather is sunny today                                0.268452
I love this new phone                                     0.240396
The meeting starts at noon                                1.000000
She feels tired and stressed                              0.234568
The package arrived late                                  0.286765
He won the tennis match                                   0.241395
This restaurant is amazing                                0.292593
The service was terrible                                  0.234370
```

```
I am learning natural language processing                0.244582
The movie was boring                                     0.270198
```