In [11]:
```
import pickle
with open('processed_text.pkl', 'rb') as file:
    processed_docs = pickle.load(file)

with open("paras.pkl", "rb") as file:  # "rb" 表示以二进制读取模式打开文件
    loaded_paras = pickle.load(file)
```

In [12]:
```
from sklearn.feature_extraction.text import TfidfVectorizer

# 导入自然语言处理工具包
import nltk
from nltk.stem import PorterStemmer  # 导入词干提取器
from nltk.stem import WordNetLemmatizer  # 导入词形还原工具
from nltk.corpus import words, stopwords, names  # 导入单词、停用词和名称库

# 下载相关数据集（如未安装）
nltk.download('words')
nltk.download('stopwords')
nltk.download('wordnet')
nltk.download('punkt')

import numpy as np
from sklearn.decomposition import TruncatedSVD
from sklearn.metrics.pairwise import cosine_similarity

from gensim.models import Word2Vec  # 导入 Word2Vec 模型
from nltk.tokenize import word_tokenize  # 导入分词工具
```

```
[nltk_data] Downloading package words to
[nltk_data]     C:\Users\sangz\AppData\Roaming\nltk_data...
[nltk_data]   Package words is already up-to-date!
[nltk_data] Downloading package stopwords to
[nltk_data]     C:\Users\sangz\AppData\Roaming\nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
[nltk_data] Downloading package wordnet to
[nltk_data]     C:\Users\sangz\AppData\Roaming\nltk_data...
[nltk_data]   Package wordnet is already up-to-date!
[nltk_data] Downloading package punkt to
[nltk_data]     C:\Users\sangz\AppData\Roaming\nltk_data...
[nltk_data]   Package punkt is already up-to-date!
```

In [13]:
```
from FlagEmbedding import FlagModel
```

In [14]:
```
from processingfunction import preprocess_paragraphs
from vectorizefunctionses import generate_tfidf_matrix
from vectorizefunctionses import get_top_similar_texts
from vectorizefunctionses import find_most_similar_document_with_reduction
from vectorizefunctionses import process_word2vec_model
from vectorizefunctionses import find_most_relevant_paragraph
```

In [16]:
```
def choose_and_call_function():
    """
    允许用户手动选择调用哪个函数，并动态传入参数。

    用户可以通过输入函数名称和对应参数来调用特定的函数。
    """
```

```
# 提供可供选择的函数列表
print("Function List: ")
print("1. get_top_similar_texts by tfidf")  # TF-IDF 检索相关文本的函数
print("2. find_most_similar_document_with_reduction by SVD")  # SVD
print("3. process_word2vec_model")  # Word2Vec 处理词向量的函数
print("4. find_most_relevant_paragraph by FlagEmbedding")  # FlagEmbedding 重

# 用户输入选择
choice = input("Choose a function: ")

# 根据选择动态调用对应的函数
if choice == "get_top_similar_texts by tfidf":
    # 初始化 TF-IDF 向量化器
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform([doc["paragraph"].lower() for doc

    # 调用函数
    query = input("query: ")
    results, top_terms, query_vector = get_top_similar_texts(query, vectorize

    # 打印结果
    print("Top Terms:")
    print(", ".join(top_terms))  # 打印 top terms 列表

    print("\nTop Similar Documents:")
    for result in results:
        print(f"Rank {result['rank']}: Document Index: {result['index']}, Sim
        print(f"Text: {result['text']}")
        print("-" * 50)

elif choice == "find_most_similar_document_with_reduction by SVD":
    # 初始化 TF-IDF 向量化器
    vectorizer = TfidfVectorizer()
    tfidf_matrix = vectorizer.fit_transform([doc["paragraph"].lower() for do

    # 调用函数
    query = input("query: ")
    results, top_terms, query_vector = get_top_similar_texts(query, vectoriz

    # 调用函数
    nearest_neighbor_index, similarity_score, nearest_document = find_most_s
        tfidf_matrix=tfidf_matrix,
        query_vector=query_vector,
        loaded_paras=loaded_paras,
        n_components=100
    )

    # 打印结果
    print(f"Most similar document index: {nearest_neighbor_index}, Similarit
    print("Most similar document content:")
    print(nearest_document)

elif choice == "process_word2vec_model":
    # 目标词
    target_word = input("query: ")

    # 调用函数
    result = process_word2vec_model(loaded_paras, target_word)

    # 打印结果
```

```
        if 'error' in result:
            print(result['error'])
        else:
            print(f"Vector of '{target_word}': {result['word_vector']}")
            print(f"Most similar to '{target_word}': {result['similar_words']}")

    elif choice == "find_most_relevant_paragraph by FlagEmbedding":
        # 示例使用
        query = input("query: ")
        result = find_most_relevant_paragraph(query, loaded_paras)

        print(f"Most relevant paragraph index: {result['index']}")
        print(f"Most relevant paragraph: {result['paragraph']}")

    else:
        print("Invalid selection!")
```

In [10]:
```
# 示例：手动调用
choose_and_call_function()
```

```
Function List:
1. get_top_similar_texts by tfidf
2. find_most_similar_document_with_reduction by SVD
3. process_word2vec_model
4. find_most_relevant_paragraph by FlagEmbedding
```

Top Terms:
second, hand, shopping, online, products, luxury, utilize, amp, guiot, acquisitio
n, roux, motivations, buying, distinct, defined, exchange, methods, theory, appro
priate, context

Top Similar Documents:
Rank 1: Document Index: 12328, Similarity Score: 0.5031925628516417
Text: {'paragraph': ' Thus, U&G theory is appropriate to utilize in the conte
xt of online second-hand luxury shopping. 2.3. Motivations for buying online seco
nd-hand luxury fashion products Second-hand shopping is defined as “the acquisiti
on of second-hand objects through methods and places of exchange that are general
ly distinct from those for new products” (Guiot & Roux, 2010, p.', 'nr': 52,
'bookID': 0}

Rank 2: Document Index: 12359, Similarity Score: 0.4980036695638522
Text: {'paragraph': ' Guiot and Roux (2010) found that second-hand consumers valu
e fashion authenticity and vintage uniqueness. Ferraro et al. (2016) found that f
ashionability plays a significant role in second-hand consumption, and consumers
who are conscious about fashion view second-hand clothing as authentic and uniqu
e.', 'nr': 83, 'bookID': 0}

Rank 3: Document Index: 12278, Similarity Score: 0.4857751519442475
Text: {'paragraph': ' Many luxury fashion retailers have been pursuing ways to ge
t involved in the second-hand goods market. However, little is known about what d
rives consumers to shop at online second-hand luxury fashion stores.', 'nr': 2,
'bookID': 0}

Rank 4: Document Index: 12294, Similarity Score: 0.47318531399679264
Text: {'paragraph': ' However, Stolz (2022) did not focus on online shopping for
second-hand luxury fashion products. Due to the growth of the online second-hand
luxury fashion market, further empirical investigation is needed to reveal young
adult consumers’ motivational factors for recommending and purchasing second-hand
luxury fashion products through online channels.', 'nr': 18, 'bookID': 0}

Rank 5: Document Index: 12330, Similarity Score: 0.46676079126751724
Text: {'paragraph': 'e. hedonic motivation), and need for a unique fashion style
(i.e. fashion motivation). These studies focused on non-luxury second-hand fashio
n products. A few studies interviewed consumers to uncover their motivations for
purchasing second-hand luxury fashion products (i.', 'nr': 54, 'bookID': 0}

Rank 6: Document Index: 12386, Similarity Score: 0.4367682872078624
Text: {'paragraph': ' This ensured that participants’ responses reflected their e
xperiences of purchasing second-hand luxury fashion products. Next, participants
were asked whether they have purchased and/or owned second-hand luxury fashion pr
oducts.', 'nr': 110, 'bookID': 0}

Rank 7: Document Index: 12440, Similarity Score: 0.4256175696711077
Text: {'paragraph': ' Whereas fashion consciousness heavily influences perceived
value for shopping at online second-hand luxury fashion retailers, status-seeking
motivation had no impact on perceived value for shopping at online second-hand lu
xury fashion retailers.', 'nr': 164, 'bookID': 0}

Rank 8: Document Index: 12573, Similarity Score: 0.42016105627245504
Text: {'paragraph': ' .82 • I feel that I can have more things for less money by
buying second-hand luxury fashion products. .82 • I feel I am paying a fair price
when I purchase second-hand luxury fashion products. .50 Critical motivation .',
'nr': 297, 'bookID': 0}

Rank 9: Document Index: 12581, Similarity Score: 0.4190513094989681
Text: {'paragraph': ' .89 • My willingness to buy a luxury fashion product from a

```
n online second-hand luxury fashion retailer is high. .89 • The probability that
I would consider buying a luxury fashion product from an online second-hand luxur
y fashion retailer is high.', 'nr': 305, 'bookID': 0}
--------------------------------------------------
Rank 10: Document Index: 12420, Similarity Score: 0.4190513094989681
Text: {'paragraph': '89•My willingness to buy a luxury fashion product from an on
line second-hand luxury fashion retailer is high.89•The probability that I would
consider buying a luxury fashion product from an online second-hand luxury fashio
n retailer is high.', 'nr': 144, 'bookID': 0}
--------------------------------------------------
```

In [17]:
```
# 示例：手动调用
choose_and_call_function()
```

```
Function List:
1. get_top_similar_texts by tfidf
2. find_most_similar_document_with_reduction by SVD
3. process_word2vec_model
4. find_most_relevant_paragraph by FlagEmbedding
Most similar document index: 12330, Similarity: 0.9589675699513049
Most similar document content:
{'paragraph': 'e. hedonic motivation), and need for a unique fashion style (i.e.
fashion motivation). These studies focused on non-luxury second-hand fashion prod
ucts. A few studies interviewed consumers to uncover their motivations for purcha
sing second-hand luxury fashion products (i.', 'nr': 54, 'bookID': 0}
```

In [18]:
```
# 示例：手动调用
choose_and_call_function()
```

```
Function List:
1. get_top_similar_texts by tfidf
2. find_most_similar_document_with_reduction by SVD
3. process_word2vec_model
4. find_most_relevant_paragraph by FlagEmbedding
second-hand fashion is not in the vocabulary
```

In [19]:
```
# 示例：手动调用
choose_and_call_function()
```

```
Function List:
1. get_top_similar_texts by tfidf
2. find_most_similar_document_with_reduction by SVD
3. process_word2vec_model
4. find_most_relevant_paragraph by FlagEmbedding
Vector of 'vintage': [-0.25137377  0.21102692 -0.46034026  0.21873228  0.36378497
0.06996898
  0.65643364  0.08505735 -0.5955887   0.24995829 -0.23649526 -0.5889536
  0.43841967  0.00457888  0.24210204 -0.4288165   0.9232003   0.2545735
 -0.31559324 -0.15861934  0.23208387  0.3573708   0.43159503 -0.04186258
 -0.03099635  0.24054135  0.13900119  0.3509962  -0.9777212  -0.10984261
  0.34451702 -0.3224587  -0.10536078  0.31209853 -0.7450898   0.59158534
  0.51292616  0.14233233 -0.35750026  0.10332837  0.52152467  0.27595162
 -0.4558136   0.1613489   0.5869396   0.08930022 -0.5167108  -0.24264237
  0.6068314  -0.07650535]
Most similar to 'vintage': [('persian/iranian', 0.9830060601234436), ('mesopotami
an', 0.9647058248519897), ('2:607', 0.9646598100662231)]
```

In [20]:
```
choose_and_call_function()
```

```
Function List:
1. get_top_similar_texts by tfidf
2. find_most_similar_document_with_reduction by SVD
3. process_word2vec_model
4. find_most_relevant_paragraph by FlagEmbedding
pre tokenize: 100%|██████████| 130/130 [00:02<00:00, 51.57it/s]
You're using a BertTokenizerFast tokenizer. Please note that with a fast tokenize
r, using the `__call__` method is faster than using a method to encode the text f
ollowed by a call to the `pad` method to get a padded encoding.
Inference Embeddings: 100%|██████████| 130/130 [00:28<00:00,  4.64it/s]
Most relevant paragraph index: 12359
Most relevant paragraph: {'paragraph': ' Guiot and Roux (2010) found that second-
hand consumers value fashion authenticity and vintage uniqueness. Ferraro et al.
(2016) found that fashionability plays a significant role in second-hand consumpt
ion, and consumers who are conscious about fashion view second-hand clothing as a
uthentic and unique.', 'nr': 83, 'bookID': 0}
```