

```

import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import string
import re
%matplotlib inline
import nltk
nltk.download('punkt')
nltk.download('wordnet')
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.tokenize import word_tokenize
from nltk.stem.wordnet import WordNetLemmatizer
lem = WordNetLemmatizer()
stop_words = set(stopwords.words('english'))
exclude = set(string.punctuation)
import string

```

```

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
[nltk_data] Downloading package wordnet to /root/nltk_data...
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.

```

```

!pip install python-terrier
import pyterrier as pt
if not pt.started():
    pt.init()

```

```

Requirement already satisfied: python-terrier in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pandas in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: wget in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pyjnius>=1.4.2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: matchpy in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: deprecated in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: chest in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: scipy in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: joblib in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: nptyping==1.4.4 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: more-itertools in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: ir-datasets>=0.3.2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: statsmodels in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: ir-measures>=0.3.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: dill in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pytreceval-terrier>=0.5.3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: typish>=1.7.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: beautifulsoup4>=4.4.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: inscriptis>=2.2.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: lxml>=4.5.2 in /usr/local/lib/python3.10/dist-packages

```

```
Requirement already satisfied: pyyaml>=5.3.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: trec-car-tools>=2.5.4 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: lz4>=3.1.10 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: warc3-wet>=0.2.3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: warc3-wet-clueweb09>=0.2.5 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: zlib-state>=0.1.3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: ijson>=3.1.3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pyautocorpus>=0.1.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: unlzw3>=0.2.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: cwl-eval>=1.0.10 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: heapdict in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: wrapt<2,>=1.10 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: multiset<3.0,>=2.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: python-dateutil>=2.8.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: threadpoolctl>=2.0.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: patsy>=0.5.2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: packaging>=21.3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: six in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: cbor>=1.0.0 in /usr/local/lib/python3.10/dist-packages
```

```
!pip install -U sentence-transformers
!pip install neuspell
!pip install -e neuspell/
!git clone https://github.com/neuspell/neuspell; cd neuspell
import os
os.chdir("/content/neuspell")
```

```
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: typing-extensions>=3.7.4.3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: packaging>=20.9 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: regex!=2019.12.17 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.10/dist-packages
```

```
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: jsonlines in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pytorch-pretrained-bert in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: Jinja2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: attrs>=19.2.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: boto3 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: huggingface-hub<1.0,>=0.16.4 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: pyyaml>=5.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: tokenizers<0.19,>=0.14 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: safetensors>=0.3.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: botocore<1.34.0,>=1.33.13 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: s3transfer<0.9.0,>=0.8.2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: mpmath>=0.19 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages
Obtaining file:///content/neuspell/neuspell
```

```
ERROR: file:///content/neuspell/neuspell does not appear to be a Python project
fatal: destination path 'neuspell' already exists and is not an empty directory
```

```
!pip install -r /content/neuspell/extras-requirements.txt
```

```
!python -m spacy download en_core_web_sm
```



```
Requirement already satisfied: pytorch-pretrained-bert in /usr/local/lib/python3.10/dist-packages
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packages
Collecting allennlp==1.5.0 (from neuspell==1.0.0->-r /content/neuspell/extras-requirements.txt)
Using cached allennlp-1.5.0-py3-none-any.whl (517 kB)
```

```
INFO: pip is looking at multiple versions of allennlp to determine which version is compatible with the requirements. You may want to disable all compatibility checks with --no-compatible-wheels.
```

```
ERROR: Cannot install neuspell, neuspell[elmo]==1.0.0 and neuspell[spacy]==1.0.0 because these package versions have conflicting dependencies.
```

The conflict is caused by:

neuspell[spacy] 1.0.0 depends on torch>=1.6.0

neuspell[elmo] 1.0.0 depends on torch>=1.6.0

allennlp 1.5.0 depends on torch<1.8.0 and >=1.6.0

To fix this you could try to:

1. loosen the range of package versions you've specified

```

2023-12-13 17:15:49.504420: W tensorflow/compiler/tf2tensorrt/utils/py_utils.cc:38
Collecting en-core-web-sm==3.6.0
  Downloading https://github.com/explosion/spacy-models/releases/download/en_core_web_sm-3.6.0/en_core_web_sm-3.6.0.tar.gz: 12.8/12.8 MB 21.6 MB/s eta 0:00
Requirement already satisfied: spacy<3.7.0,>=3.6.0 in /usr/local/lib/python3.10/site-packages (3.6.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.10/site-packages (3.0.11)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.10/site-packages (1.0.0)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.10/site-packages (0.28.0)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.10/site-packages (2.0.2)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.10/site-packages (3.0.2)
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.10/site-packages (8.1.8)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.10/site-packages (0.9.1)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.10/site-packages (2.4.3)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.10/site-packages (2.0.6)
Requirement already satisfied: typer<0.10.0,>=0.3.0 in /usr/local/lib/python3.10/site-packages (0.3.0)
Requirement already satisfied: pathy<0.10.0 in /usr/local/lib/python3.10/site-packages (0.10.0)
Requirement already satisfied: smart-open<7.0.0,>=5.2.1 in /usr/local/lib/python3.10/site-packages (5.2.1)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.10/site-packages (4.38.0)
Requirement already satisfied: numpy<1.15.0 in /usr/local/lib/python3.10/site-packages (1.15.0)
Requirement already satisfied: requests<3.0.0,>=2.13.0 in /usr/local/lib/python3.10/site-packages (2.13.0)
Requirement already satisfied: pydantic!=1.8,!=1.8.1,<3.0.0,>=1.7.4 in /usr/local/lib/python3.10/site-packages (1.7.4)
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/site-packages (2.11.2)
Requirement already satisfied: setuptools in /usr/local/lib/python3.10/site-packages (59.0.1)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/site-packages (23.1)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.10/site-packages (3.2.0)
Requirement already satisfied: typing-extensions>=4.2.0 in /usr/local/lib/python3.10/site-packages (4.5.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.10/site-packages (3.3.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.10/site-packages (3.4)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.10/site-packages (1.26.15)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.10/site-packages (2023.7.22)
Requirement already satisfied: blis<0.8.0,>=0.7.8 in /usr/local/lib/python3.10/site-packages (0.7.8)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.10/site-packages (0.0.1)
Requirement already satisfied: click<9.0.0,>=7.1.1 in /usr/local/lib/python3.10/site-packages (8.1.7)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.10/site-packages (2.1.2)
✓ Download and installation successful
You can now load the package via spacy.load('en_core_web_sm')

```

```
pip install neuspell[elmo]~=1.0.0
```

```

Requirement already satisfied: neuspell[elmo]~=1.0.0 in /usr/local/lib/python3.10/site-packages (1.0.0)
Requirement already satisfied: transformers in /usr/local/lib/python3.10/site-packages (4.31.0)
Requirement already satisfied: tqdm in /usr/local/lib/python3.10/site-packages (4.38.0)
Requirement already satisfied: torch>=1.6.0 in /usr/local/lib/python3.10/site-packages (1.13.1)
Requirement already satisfied: numpy in /usr/local/lib/python3.10/site-packages (1.24.2)
Requirement already satisfied: jsonlines in /usr/local/lib/python3.10/site-packages (3.2.0)
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/site-packages (0.1.99)
Requirement already satisfied: pytorch-pretrained-bert in /usr/local/lib/python3.10/site-packages (0.0.2)
Collecting allennlp==1.5.0 (from neuspell[elmo]~=1.0.0)
  Using cached allennlp-1.5.0-py3-none-any.whl (517 kB)
INFO: pip is looking at multiple versions of allennlp to determine which version is compatible with the requirements. This process will find the best available version in the current project's environment.
Collecting neuspell[elmo]~=1.0.0
  Using cached neuspell-1.0.0-py3-none-any.whl (184 kB)
ERROR: Cannot install neuspell and neuspell[elmo]==1.0.0 because these package versions have conflicting dependencies.

The conflict is caused by:
  neuspell[elmo] 1.0.0 depends on torch>=1.6.0
  allennlp 1.5.0 depends on torch<1.8.0 and >=1.6.0

```

To fix this you could try to:

1. loosen the range of package versions you've specified

2. remove package versions to allow pip attempt to solve the dependency confl:

ERROR: ResolutionImpossible: for help visit <https://pip.pypa.io/en/latest/top>

```
!pip install -e .[elmo]
```

```
!pip install -e .[spacy]
```

```
Requirement already satisfied: torch>=1.6.0 in /usr/local/lib/python3.10/dist-
Requirement already satisfied: numpy in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: jsonlines in /usr/local/lib/python3.10/dist-pa
Requirement already satisfied: sentencepiece in /usr/local/lib/python3.10/dis
Requirement already satisfied: pytorch_pretrained_bert in /usr/local/lib/pytho
Requirement already satisfied: spacy in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: filelock in /usr/local/lib/python3.10/dist-pacl
Requirement already satisfied: typing-extensions in /usr/local/lib/python3.10,
Requirement already satisfied: sympy in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: networkx in /usr/local/lib/python3.10/dist-pacl
Requirement already satisfied: jinja2 in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: fsspec in /usr/local/lib/python3.10/dist-packa
Requirement already satisfied: triton==2.1.0 in /usr/local/lib/python3.10/dis
Requirement already satisfied: attrs>=19.2.0 in /usr/local/lib/python3.10/dis
Requirement already satisfied: boto3 in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: requests in /usr/local/lib/python3.10/dist-pacl
Requirement already satisfied: regex in /usr/local/lib/python3.10/dist-packag
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/|
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/|
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/py
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python
Requirement already satisfied: thinc<8.2.0,>=8.1.8 in /usr/local/lib/python3.
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.
```

```

Attempting uninstall: neuspell
Found existing installation: neuspell 1.0.0
Uninstalling neuspell-1.0.0:
  Successfully uninstalled neuspell-1.0.0
Running setup.py develop for neuspell
Successfully installed neuspell-1.0.0

```

```

from google.colab import drive
drive.mount('/content/drive')

```

Drive already mounted at /content/drive; to attempt to forcibly remount, call

```

#Unzipping the multi-linguistic packages
!wget https://storage.googleapis.com/bert_models/2018_11_23/multi_cased_L-12_H-7
!unzip *.zip

--2023-12-13 17:15:21-- https://storage.googleapis.com/bert_models/2018_11_23/
Resolving storage.googleapis.com (storage.googleapis.com)... 108.177.11.207, :
Connecting to storage.googleapis.com (storage.googleapis.com)|108.177.11.207|
HTTP request sent, awaiting response... 403 Forbidden
2023-12-13 17:15:21 ERROR 403: Forbidden.

--2023-12-13 17:15:21-- http://l-12_h-768_a-12.zip/
Resolving l-12_h-768_a-12.zip (l-12_h-768_a-12.zip)... failed: Name or service
wget: unable to resolve host address 'l-12_h-768_a-12.zip'
unzip: cannot find or open *.zip, *.zip.zip or *.zip.ZIP.

No zipfiles found.

```

```

import pandas as pd

# Replace 'your_csv_url' with the actual URL of your CSV file
csv_url = '/flipkart_com-ecommerce_sample.csv'

# Read the CSV file
df = pd.read_csv(csv_url)

# Display the DataFrame
df.head()

```

	uniq_id	crawl_timestamp	product_u
0	c2d766ca982eca8304150849735ffef9	2016-03-25 22:59:23 +0000	http://www.flipkart.com/alisha-so women-s-
1	7f7036a6d550aaa89d34c77bd39a5e48	2016-03-25 22:59:23 +0000	http://www.flipkart.com/fabhomedec fabric-d

```
from google.colab import drive
drive.mount('/content/gdrive')
```

Mounted at /content/gdrive

```
# Save a DataFrame to a file in the Colab environment
df.to_csv('/flipkart_com-ecommerce_sample.csv', index=False)
```

```
# To download a file in Colab
from google.colab import files
files.download('/flipkart_com-ecommerce_sample.csv')
```

```
df['product_category_tree']=df['product_category_tree'].map(lambda x:x. strip('[]
df['product_category_tree']=df['product_category_tree'].map(lambda x:x. strip('"'
df['product_category_tree']=df['product_category_tree'].map(lambda x:x. split('>>
```

```
#next, drop unwanted columns.
df.drop(['crawl_timestamp','product_url',
        'image',"retail_price","discounted_price",
        "is_FK_Advantage_product","product_rating","overall_rating","product_spe
```

```
#Then, drop duplicate products.
uniq_prod=df.copy()
uniq_prod.drop_duplicates(subset ="product_name", keep = "first", inplace = True)
#Remove stop words and punctuations and then perform tokenization and lemmatizati
def filter_keywords(doc):
    doc=doc.lower()
    stop_free = " ".join([i for i in doc.split() if i not in stop_words])
    punc_free = "".join(ch for ch in stop_free if ch not in exclude)
    word_tokens = word_tokenize(punc_free)
    filtered_sentence = [(lem.lemmatize(w, "v")) for w in word_tokens]
    return filtered_sentence
#Apply the filter_keywords function to selected columns to obtain the keywords fo
uniq_prod['product'] = uniq_prod['product_name'].apply(filter_keywords)
uniq_prod['description'] = uniq_prod['description'].astype("str"). apply(filter_k
uniq_prod['brand'] = uniq_prod['brand'].astype("str").apply(filter_keywords)
```

```
#Combine all the keywords for each product.
uniq_prod["keywords"]=uniq_prod['product']+uniq_prod['brand']+ df['product_catego
uniq_prod["keywords"] = uniq_prod["keywords"].apply(lambda x: ' '.join(x))

#Creating a 'docno' column, which gives recommendations.
uniq_prod['docno']=uniq_prod['product_name']

#Drop unwanted columns.
uniq_prod.drop(['product','brand','pid','product_name'],axis=1,inplace=True)
uniq_prod.head()
```

	uniq_id	product_category_tree	description	keywo
0	c2d766ca982eca8304150849735ffef9	[Clothing , Women's Clothing , Lingerie, Sle...	[key, feature, alisha, solid, womens, cycle, s...	alisha : womens c short al Cloth
1	7f7036a6d550aaa89d34c77bd39a5e48	[Furniture , Living Room Furniture , Sofa Be...	[fabhomedecor, fabric, double, sofa, bed, fini...	fabhomed fabric do sofa fabhomed
2	f449ec65dcbc041b6ae5e6a32717d01b	[Footwear , Women's Footwear , Ballerinas , ...	[key, feature, aw, belly, sandals, wedge, heel...	aw bell Foot Wom Foot B
4	bc940ea42ee6bef5ac7cea3fb5cfbee7	[Pet Supplies , Grooming , Skin & Coat Care ...	[specifications, sicons, purpose, arnica, dog,...	sic purp arnica shar sicons Pei
5	c2a17313954882c1dba461863e98adf2	[Eternal Gandhi Super Series Crystal Paper Wei...	[key, feature, eternal, gandhi, super, series,...	eternal ga super se crystal p we

Building the Search Engine

```
!rm -rf /content/iter_index_porter
pd_indexer = pt.DFIndexer("/content/pd_index")
indexref = pd_indexer.index(uniq_prod["keywords"], uniq_prod["docno"])

spellcheck = BertChecker()
spellcheck.from_pretrained(
    ckpt_path=f"/content/multi_cased_L-12_H-768_A-12" # "<folder where the model
)
```



```
loading vocab from path:/content/neuspell/neuspell/../data/checkpoints/subwor
initializing model
Number of parameters in the model: 185211810
Loading model params from checkpoint dir: /content/neuspell/neuspell/../data/
```

```
-----
UnpicklingError                                Traceback (most recent call last)
<ipython-input-30-2b842f628aaf> in <cell line: 2>()
      1 spellcheck = BertChecker()
----> 2 spellcheck.from_pretrained(
      3     ckpt_path=f"/content/multi_cased_L-12_H-768_A-12" # "<folder
where the model is saved>"
      4 )
```

```
----- 5 frames -----
/usr/local/lib/python3.10/dist-packages/torch/serialization.py in
_legacy_load(f, map_location, pickle_module, **pickle_load_args)
    1244         "functionality.")
    1245
-> 1246     magic_number = pickle_module.load(f, **pickle_load_args)
    1247     if magic_number != MAGIC_NUMBER:
    1248         raise RuntimeError("Invalid magic number; corrupt file?")
```

UnpicklingError: invalid load key, '<'.

EXPLAIN ERROR

```
!pip install torch==1.13.1
```

```
spellcheck = BertChecker()
spellcheck.from_pretrained(
    ckpt_path=f"/content/multi_cased_L-12_H-768_A-12" # "<folder where the model
)
X=input("Search Engine:")
query=spellcheck.correct(X)
print(query)
```

```
prod_ret = pt.BatchRetrieve(indexref, wmodel='TF_IDF',
properties={'termpipelines': 'Stopwords'})
pr=prod_ret.compile()
output=pr.search(query)
docno=list(output['docno'])
transform=model.encode(docno)
```

```

from neuspell import BertChecker

# Initialize spell checker
spellcheck = BertChecker()

# Load the pre-trained model and vocabulary

# Check if the model and vocab are ready
spellcheck.is_model_ready()

# Now you can use the correct method
X = input("Search Engine:")
query = spellcheck.correct(X)
print(query)

```

model & vocab must be loaded first

```

-----
AssertionError                                Traceback (most recent call last)
<ipython-input-34-8312cd0be1d7> in <cell line: 9>()
      7
      8 # Check if the model and vocab are ready
----> 9 spellcheck.is_model_ready()
     10
     11 # Now you can use the correct method

/content/neuspell/neuspell/corrector.py in is_model_ready(self)
      71
      72     def is_model_ready(self):
----> 73         assert not (self.model is None or self.vocab is None),
print("model & vocab must be loaded first")
      74
      75     @property

```

Start coding or [generate](#) with AI.

EXPLAIN ERROR