

```

//DOVANKHAI_CNTT4_10050090140
from pybloom_live import BloomFilter
import nltk
nltk.download('punkt')
# Tải tập từ điển từ thư viện nltk
nltk.download('words')

# Đường dẫn đến file chứa văn bản
file_path = "wordcount.txt"

# Đọc nội dung từ file
with open(file_path, 'r') as file:
    text = file.read()

# Tokenize văn bản thành danh sách các từ
words = nltk.word_tokenize(text)

# Khởi tạo Bloom Filter với dung lượng và tỷ lệ lỗi cho trước
bloom_filter = BloomFilter(capacity=100000, error_rate=0.001)

# Thêm từng từ vào Bloom Filter
for word in words:
    bloom_filter.add(word)

# Danh sách các từ cần kiểm tra
words_to_check = ["of", "example", "some", "words", "to", "check"]

# Kiểm tra từng từ và in kết quả
for check in words_to_check:
    if check in bloom_filter:
        print(f"Từ '{check}' Có trong văn bản.")
    else:
        print(f"Từ '{check}' Không có trong văn bản.")

```

B1: thêm thư viện từ bên thứ 3 vào

```

: //DOVANKHAI_CNTT4_10050090140
  from pybloom_live import BloomFilter
  import nltk

```

B2: tải thư viện và thêm đường dẫn

```

nltk.download('punkt')
# Tải tập từ điển từ thư viện nltk
nltk.download('words')

# Đường dẫn đến file chứa văn bản
file_path = "wordcount.txt"

```

B3: Cài đặt code

```

# Đọc nội dung từ file
with open(file_path, 'r') as file:
    text = file.read()

# Tokenize văn bản thành danh sách các từ
words = nltk.word_tokenize(text)

# Khởi tạo Bloom Filter với dung lượng và tỷ lệ lỗi cho trước
bloom_filter = BloomFilter(capacity=100000, error_rate=0.001)

# Thêm từng từ vào Bloom Filter
for word in words:
    bloom_filter.add(word)

# Danh sách các từ cần kiểm tra
words_to_check = ["of", "example", "some", "words", "to", "check"]

# Kiểm tra từng từ và in kết quả
for check in words_to_check:
    if check in bloom_filter:
        print(f"Từ '{check}' Có trong văn bản.")
    else:
        print(f"Từ '{check}' Không có trong văn bản.")

```

B4: chụp hình kết quả

```

[nltk_data] Downloading package punkt to /home/khaicntt4/nltk_data...
[nltk_data]   Unzipping tokenizers/punkt.zip.
Từ 'of' Có trong văn bản.
Từ 'example' Có trong văn bản.
Từ 'some' Không có trong văn bản.
Từ 'words' Không có trong văn bản.
Từ 'to' Có trong văn bản.
Từ 'check' Có trong văn bản.

[nltk_data] Downloading package words to /home/khaicntt4/nltk_data...
[nltk_data]   Package words is already up-to-date!

```