

LAB 1,2 đã cài hadoop nên em skip

```
khaicntt4@khaicntt4-virtual-machine: ~/hadoop-3.3.6$ start-all.sh
WARNING: Attempting to start all Apache Hadoop daemons as khaicntt4 in 10 seconds.
WARNING: This is not a recommended production deployment configuration.
WARNING: Use CTRL-C to abort.
Starting namenodes on [localhost]
Starting datanodes
Starting secondary namenodes [khaicntt4-virtual-machine]
Starting resourcemanager
Starting nodemanagers
```

LAB 3,4 WORDCOUNT + STOPWORD CODE

B1:

```
: import nltk
  from nltk.corpus import stopwords
  from nltk.tokenize import word_tokenize
```

B2: CHÈN TỪ KHOÁ CẦN LỌC

text = "decision trees are a very effective method of supervised learning.it aims is the partition of a dataset into groups as homogeneous as possible in terms of the variable to be predicted it takes as input a set of classified data, and outpu

B3: CÀI ĐẶT CODE CỦA STOPWORD

```

tokens = word_tokenize(text)

tokens = [word.lower() for word in tokens]
#layds tu trong tieng anh
stop_words = set(stopwords.words('english'))

stopwords_count = {}
for word in tokens:
    if word in stop_words:
        if word in stopwords_count:
            stopwords_count[word] += 1
        else:
            stopwords_count[word] = 1

for word, count in stopwords_count.items():
    print(f"{word}: {count}")
def count_word_frequency(file_path):
    word_frequency = {}
    with open(file_path, 'r') as file:
        content = file.read()
        words = content.split()
        for word in words:
            if word in word_frequency:
                word_frequency[word] += 1
            else:
                word_frequency[word] = 1
    return word_frequency

```

B4: CÀI ĐẶT CODE CỦA WORDCOUNT

```

file_path = 'wordcount.txt'
word_frequency = count_word_frequency(file_path)
for word, frequency in word_frequency.items():
    print(f"Từ '{word}' xuất hiện {frequency} lần")

```

B5: SHOW KQ CỦA STOPWORD

are: 1
a: 8
very: 1
of: 6
is: 2
the: 5
into: 1
as: 3
in: 1
to: 4
be: 1
it: 1
and: 2
that: 1
an: 1
where: 1
each: 3
all: 1
from: 1

B5: SHOW KQ CỦA WORDCOUNT

from: 4

Từ 'How' xuất hiện 1 lần
Từ 'Blockchain' xuất hiện 1 lần
Từ 'Works' xuất hiện 1 lần
Từ 'When' xuất hiện 3 lần
Từ 'a' xuất hiện 13 lần
Từ 'block' xuất hiện 10 lần
Từ 'stores' xuất hiện 1 lần
Từ 'new' xuất hiện 3 lần
Từ 'data' xuất hiện 2 lần
Từ 'it' xuất hiện 5 lần
Từ 'is' xuất hiện 4 lần
Từ 'added' xuất hiện 6 lần
Từ 'to' xuất hiện 11 lần
Từ 'the' xuất hiện 19 lần
Từ 'blockchain.' xuất hiện 4 lần
Từ 'Blockchain,' xuất hiện 1 lần
Từ 'as' xuất hiện 3 lần
Từ 'its' xuất hiện 2 lần
Từ 'name' xuất hiện 1 lần
Từ 'suggests,' xuất hiện 1 lần
Từ 'consists' xuất hiện 1 lần
Từ 'of' xuất hiện 11 lần
Từ 'multiple' xuất hiện 2 lần
Từ 'blocks' xuất hiện 1 lần
Từ 'strung' xuất hiện 1 lần
Từ 'together.' xuất hiện 1 lần
Từ 'In' xuất hiện 1 lần
Từ 'order' xuất hiện 1 lần
Từ 'for' xuất hiện 2 lần
Từ 'be' xuất hiện 8 lần
Từ 'blockchain,' xuất hiện 4 lần
Từ 'however,' xuất hiện 2 lần
Từ 'four' xuất hiện 1 lần
Từ 'things' xuất hiện 1 lần
Từ 'must' xuất hiện 7 lần
Từ 'happen' xuất hiện 1 lần
Từ 'A' xuất hiện 1 lần
Từ 'transaction' xuất hiện 9 lần
Từ 'occur.' xuất hiện 1 lần
Từ 'Let's' xuất hiện 1 lần
Từ 'continue' xuất hiện 1 lần
Từ 'with' xuất hiện 3 lần
Từ 'example' xuất hiện 1 lần
Từ 'your' xuất hiện 9 lần
Từ 'impulsive' xuất hiện 1 lần
Từ 'Amazon' xuất hiện 2 lần
Từ 'purchase.' xuất hiện 2 lần
Từ 'After' xuất hiện 3 lần
Từ 'hastily' xuất hiện 1 lần
Từ 'clicking' xuất hiện 1 lần
Từ 'through' xuất hiện 1 lần
Từ 'checkout' xuất hiện 1 lần
Từ 'prompt,' xuất hiện 1 lần
Từ 'you' xuất hiện 6 lần
Từ 'and' xuất hiện 1 lần

Từ 'things' xuất hiện 1 lần
Từ 'must' xuất hiện 7 lần
Từ 'happen' xuất hiện 1 lần
Từ 'A' xuất hiện 1 lần
Từ 'transaction' xuất hiện 9 lần
Từ 'occur.' xuất hiện 1 lần
Từ 'Let's' xuất hiện 1 lần
Từ 'continue' xuất hiện 1 lần
Từ 'with' xuất hiện 3 lần
Từ 'example' xuất hiện 1 lần
Từ 'your' xuất hiện 9 lần
Từ 'impulsive' xuất hiện 1 lần
Từ 'Amazon' xuất hiện 2 lần
Từ 'purchase.' xuất hiện 2 lần
Từ 'After' xuất hiện 3 lần
Từ 'hastily' xuất hiện 1 lần
Từ 'clicking' xuất hiện 1 lần
Từ 'through' xuất hiện 1 lần
Từ 'checkout' xuất hiện 1 lần
Từ 'prompt,' xuất hiện 1 lần
Từ 'you' xuất hiện 6 lần
Từ 'go' xuất hiện 1 lần
Từ 'against' xuất hiện 1 lần
Từ 'better' xuất hiện 1 lần
Từ 'judgment' xuất hiện 1 lần
Từ 'and' xuất hiện 4 lần
Từ 'make' xuất hiện 2 lần
Từ 'As' xuất hiện 1 lần
Từ 'we' xuất hiện 1 lần
Từ 'discussed' xuất hiện 1 lần
Từ 'above,' xuất hiện 1 lần
Từ 'in' xuất hiện 7 lần
Từ 'many' xuất hiện 1 lần
Từ 'cases' xuất hiện 1 lần
Từ 'will' xuất hiện 4 lần
Từ 'group' xuất hiện 1 lần
Từ 'together' xuất hiện 1 lần
Từ 'potentially' xuất hiện 1 lần
Từ 'thousands' xuất hiện 1 lần
Từ 'transactions,' xuất hiện 1 lần
Từ 'so' xuất hiện 1 lần
Từ 'purchase' xuất hiện 2 lần
Từ 'packaged' xuất hiện 1 lần
Từ 'along' xuất hiện 2 lần
Từ 'other' xuất hiện 2 lần
Từ 'users'' xuất hiện 1 lần
Từ 'information' xuất hiện 2 lần
Từ 'well.' xuất hiện 1 lần
Từ 'That' xuất hiện 4 lần
Từ 'verified.' xuất hiện 2 lần
Từ 'making' xuất hiện 1 lần
Từ 'that' xuất hiện 6 lần
Từ 'purchase,' xuất hiện 2 lần
Từ 'With' xuất hiện 2 lần
Từ 'public' xuất hiện 1 lần
Từ 'records' xuất hiện 1 lần
Từ 'information' xuất hiện 1 lần

Từ 'thousands' xuất hiện 1 lần
Từ 'transactions,' xuất hiện 1 lần
Từ 'so' xuất hiện 1 lần
Từ 'purchase' xuất hiện 2 lần
Từ 'packaged' xuất hiện 1 lần
Từ 'along' xuất hiện 2 lần
Từ 'other' xuất hiện 2 lần
Từ 'users'' xuất hiện 1 lần
Từ 'information' xuất hiện 2 lần
Từ 'well.' xuất hiện 1 lần
Từ 'That' xuất hiện 4 lần
Từ 'verified.' xuất hiện 2 lần
Từ 'making' xuất hiện 1 lần
Từ 'that' xuất hiện 6 lần
Từ 'purchase,' xuất hiện 2 lần
Từ 'With' xuất hiện 2 lần
Từ 'public' xuất hiện 1 lần
Từ 'records' xuất hiện 1 lần
Từ 'information,' xuất hiện 1 lần
Từ 'like' xuất hiện 2 lần
Từ 'Securities' xuất hiện 1 lần
Từ 'Exchange' xuất hiện 1 lần
Từ 'Commission,' xuất hiện 1 lần
Từ 'Wikipedia,' xuất hiện 1 lần
Từ 'or' xuất hiện 2 lần
Từ 'local' xuất hiện 1 lần
Từ 'library,' xuất hiện 1 lần
Từ 'there's' xuất hiện 1 lần
Từ 'someone' xuất hiện 1 lần
Từ 'charge' xuất hiện 1 lần
Từ 'vetting' xuất hiện 1 lần
Từ 'entries.' xuất hiện 1 lần
Từ 'job' xuất hiện 1 lần
Từ 'left' xuất hiện 1 lần
Từ 'up' xuất hiện 1 lần
Từ 'network' xuất hiện 2 lần
Từ 'computers.' xuất hiện 1 lần
Từ 'from' xuất hiện 1 lần
Từ 'Amazon,' xuất hiện 1 lần
Từ 'computers' xuất hiện 1 lần
Từ 'rushes' xuất hiện 1 lần
Từ 'check' xuất hiện 1 lần
Từ 'happened' xuất hiện 1 lần
Từ 'way' xuất hiện 1 lần
Từ 'said' xuất hiện 1 lần
Từ 'did.' xuất hiện 1 lần
Từ 'is,' xuất hiện 1 lần
Từ 'they' xuất hiện 1 lần
Từ 'confirm' xuất hiện 1 lần
Từ 'details' xuất hiện 1 lần
Từ 'including' xuất hiện 1 lần
Từ 'transaction's' xuất hiện 2 lần
Từ 'time,' xuất hiện 1 lần
Từ 'dollar' xuất hiện 2 lần
Từ 'amount,' xuất hiện 2 lần
Từ 'participants.' xuất hiện 1 lần
Từ '(More' xuất hiện 1 lần

Từ '(More' xuất hiện 1 lần
Từ 'on' xuất hiện 1 lần
Từ 'how' xuất hiện 1 lần
Từ 'this' xuất hiện 1 lần
Từ 'happens' xuất hiện 1 lần
Từ 'second.)' xuất hiện 1 lần
Từ 'stored' xuất hiện 2 lần
Từ 'block.' xuất hiện 2 lần
Từ 'has' xuất hiện 1 lần
Từ 'been' xuất hiện 2 lần
Từ 'verified' xuất hiện 1 lần
Từ 'accurate,' xuất hiện 1 lần
Từ 'gets' xuất hiện 1 lần
Từ 'green' xuất hiện 1 lần
Từ 'light.' xuất hiện 1 lần
Từ 'The' xuất hiện 2 lần
Từ 'digital' xuất hiện 2 lần
Từ 'signature,' xuất hiện 1 lần
Từ 'Amazon's' xuất hiện 1 lần
Từ 'signature' xuất hiện 1 lần
Từ 'are' xuất hiện 1 lần
Từ 'all' xuất hiện 2 lần
Từ 'There,' xuất hiện 1 lần
Từ 'likely' xuất hiện 1 lần
Từ 'join' xuất hiện 1 lần
Từ 'hundreds,' xuất hiện 1 lần
Từ 'thousands,' xuất hiện 1 lần
Từ 'others' xuất hiện 1 lần
Từ 'it.' xuất hiện 1 lần
Từ 'given' xuất hiện 3 lần
Từ 'hash.' xuất hiện 2 lần
Từ 'Not' xuất hiện 1 lần
Từ 'unlike' xuất hiện 1 lần
Từ 'an' xuất hiện 1 lần
Từ 'angel' xuất hiện 1 lần
Từ 'earning' xuất hiện 1 lần
Từ 'wings,' xuất hiện 1 lần
Từ 'once' xuất hiện 1 lần
Từ 'block's' xuất hiện 1 lần
Từ 'transactions' xuất hiện 1 lần
Từ 'have' xuất hiện 2 lần
Từ 'verified,' xuất hiện 1 lần
Từ 'unique,' xuất hiện 1 lần
Từ 'identifying' xuất hiện 1 lần
Từ 'code' xuất hiện 1 lần
Từ 'called' xuất hiện 1 lần
Từ 'also' xuất hiện 1 lần
Từ 'hash' xuất hiện 1 lần
Từ 'most' xuất hiện 1 lần
Từ 'recent' xuất hiện 1 lần
Từ 'Once' xuất hiện 1 lần
Từ 'hashed,' xuất hiện 1 lần
Từ 'can' xuất hiện 1 lần
Từ 'becomes' xuất hiện 1 lần
Từ 'publicly' xuất hiện 1 lần
Từ 'available' xuất hiện 1 lần
Từ 'anyone' xuất hiện 1 lần
Từ 'view—even' xuất hiện 1 lần
Từ 'view—even' xuất hiện 1 lần

```
Từ 'called' xuất hiện 1 lần
Từ 'also' xuất hiện 1 lần
Từ 'hash' xuất hiện 1 lần
Từ 'most' xuất hiện 1 lần
Từ 'recent' xuất hiện 1 lần
Từ 'Once' xuất hiện 1 lần
Từ 'hashed,' xuất hiện 1 lần
Từ 'can' xuất hiện 1 lần
Từ 'becomes' xuất hiện 1 lần
Từ 'publicly' xuất hiện 1 lần
Từ 'available' xuất hiện 1 lần
Từ 'anyone' xuất hiện 1 lần
Từ 'view-even' xuất hiện 1 lần
Từ 'you.' xuất hiện 1 lần
Từ 'If' xuất hiện 1 lần
Từ 'take' xuất hiện 1 lần
Từ 'look' xuất hiện 1 lần
Từ 'at' xuất hiện 1 lần
Từ 'Bitcoin's' xuất hiện 1 lần
Từ 'see' xuất hiện 1 lần
Từ 'access' xuất hiện 1 lần
Từ 'data,' xuất hiện 1 lần
Từ 'about' xuất hiện 1 lần
Từ 'when' xuất hiện 1 lần
Từ '("Time"),' xuất hiện 1 lần
Từ 'where' xuất hiện 1 lần
Từ '("Height"),' xuất hiện 1 lần
Từ 'by' xuất hiện 1 lần
Từ 'who' xuất hiện 1 lần
Từ '("Relayed" xuất hiện 1 lần
Từ 'By')' xuất hiện 1 lần
Từ 'was' xuất hiện 1 lần

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

[ ]:
```