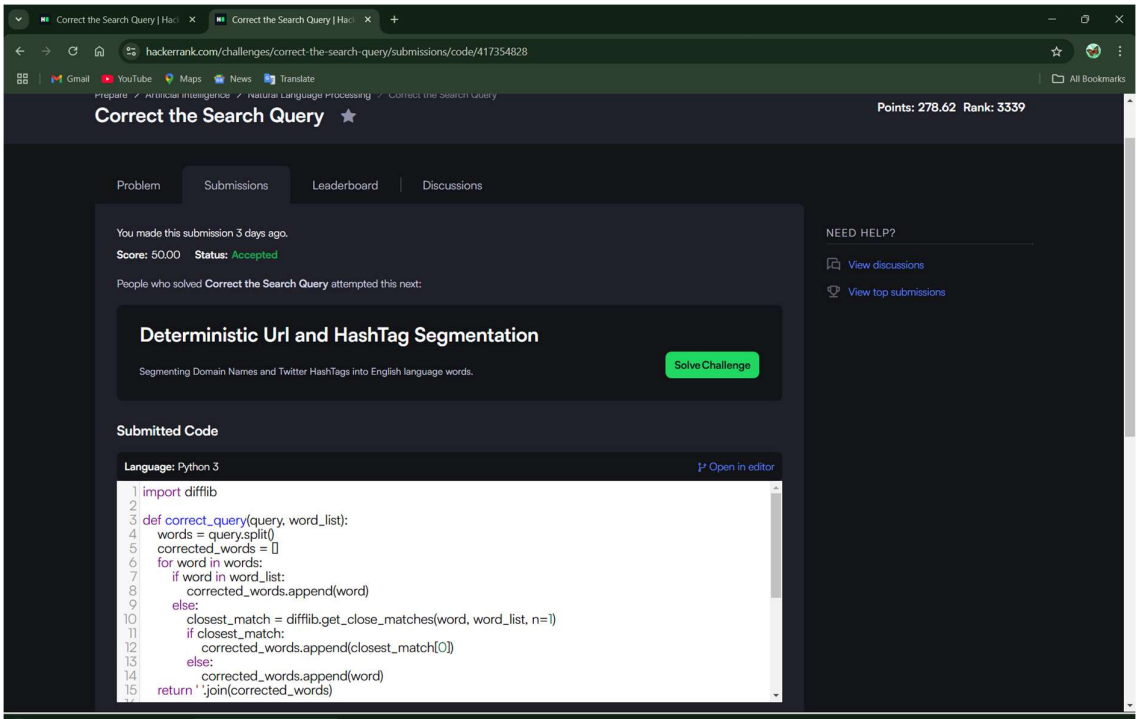
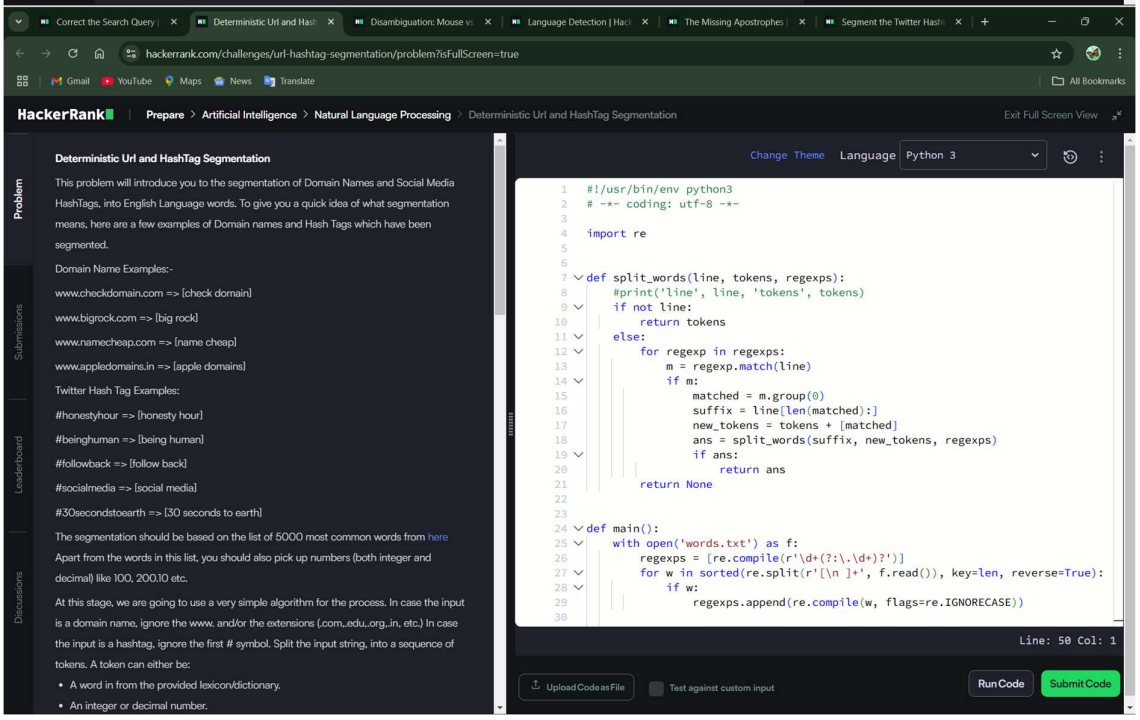
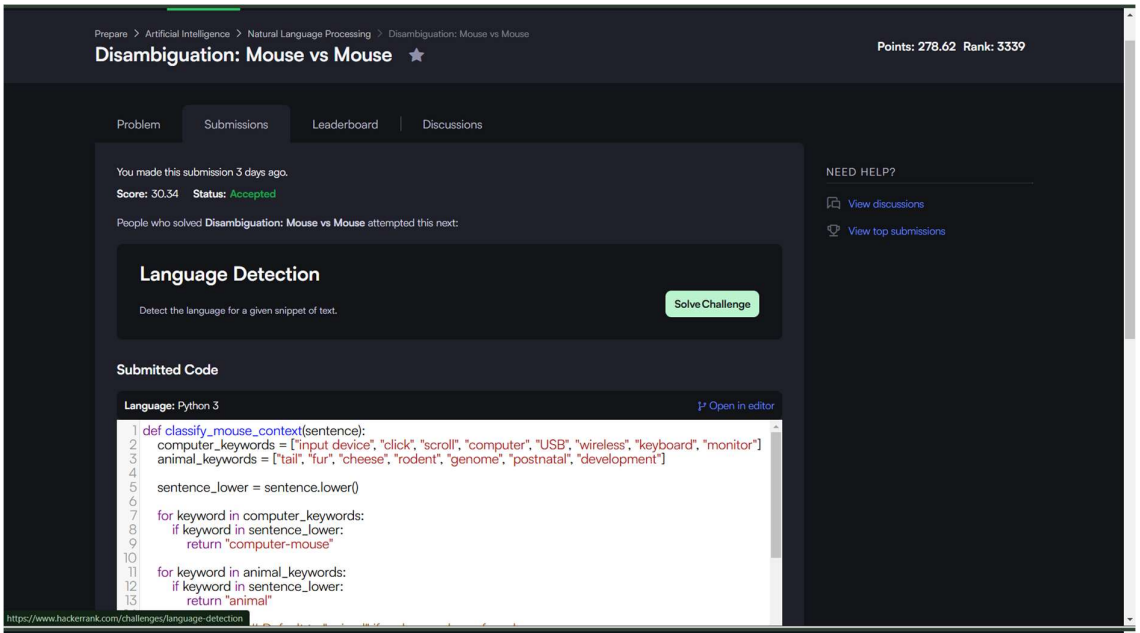
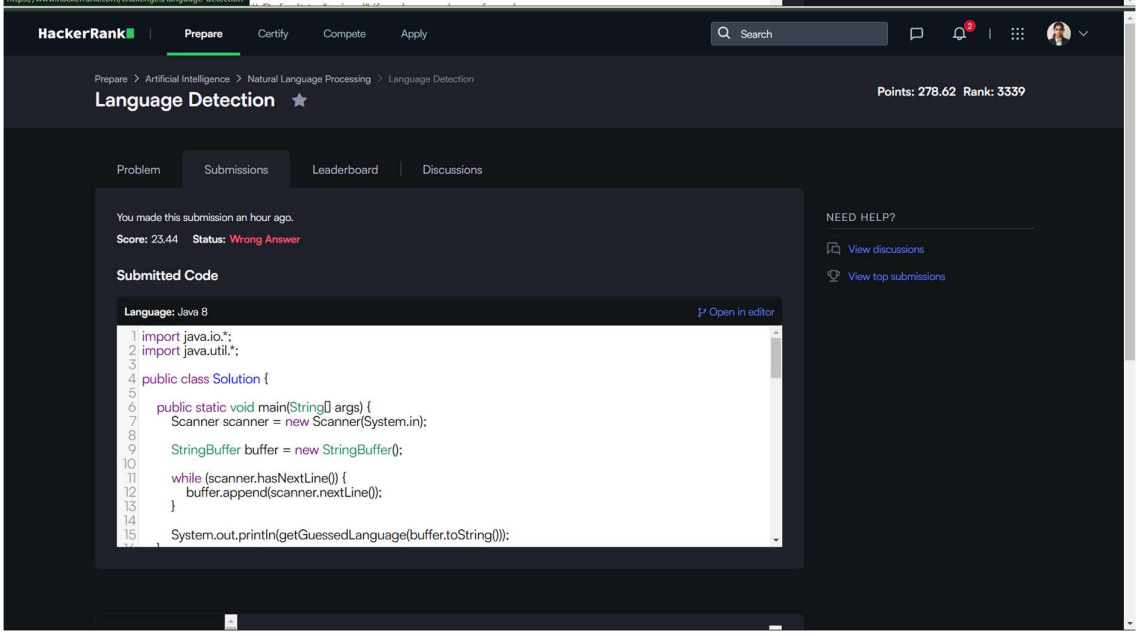
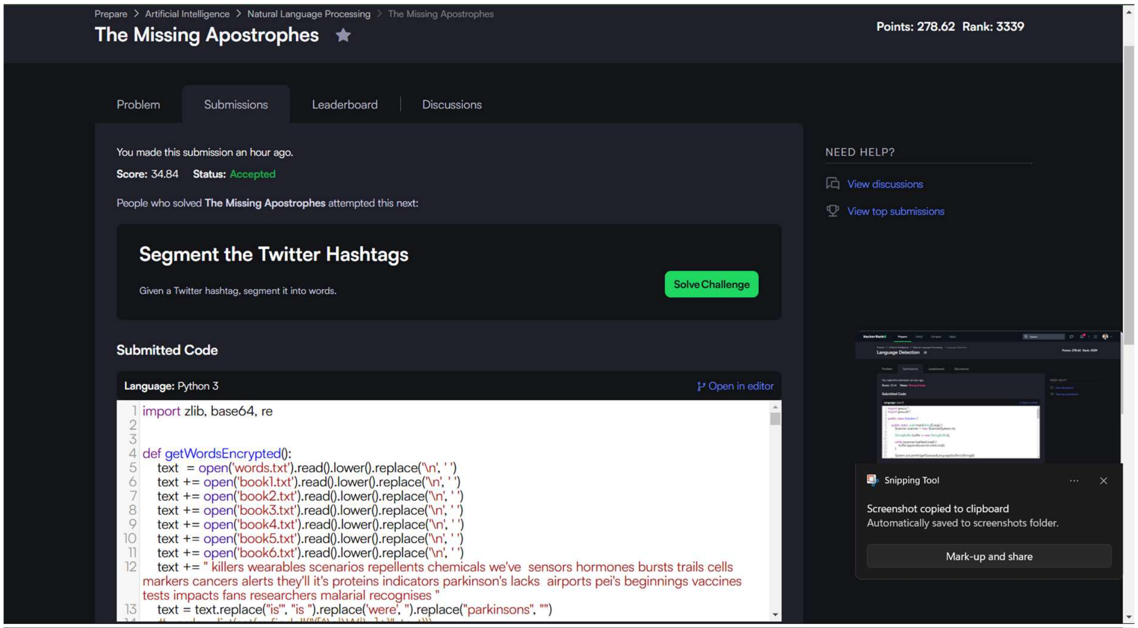
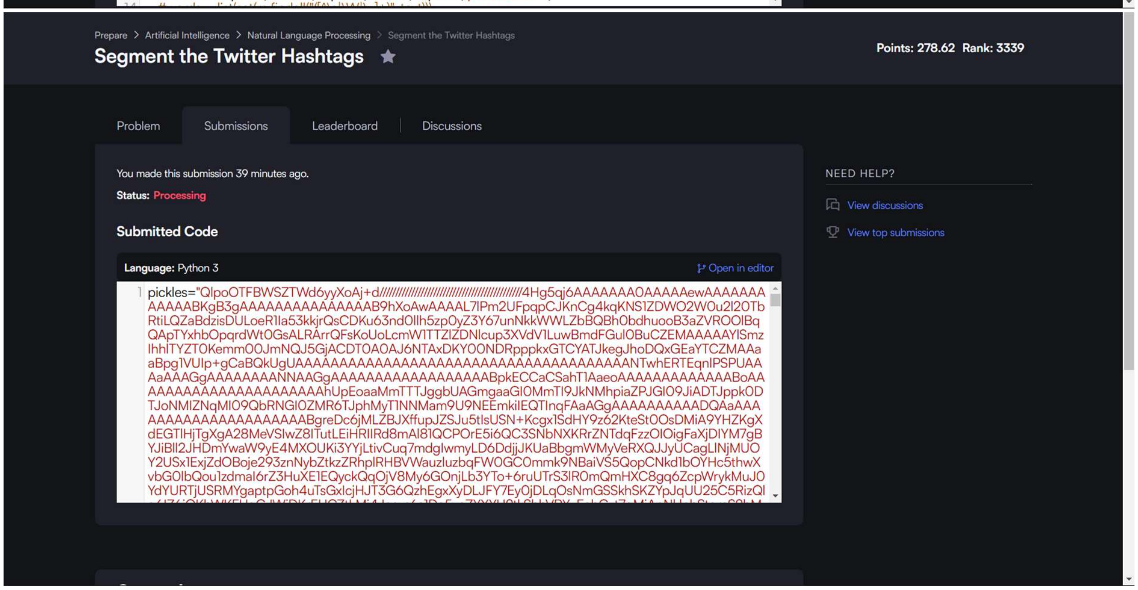

Correct the Search Query
Points: 278.62 Rank: 3339
Problem
Submissions
Leaderboard
Discussions
You made this submission 3 days ago.
Score: 50.00 Status: Accepted
People who solved Correct the Search Query attempted this next:
Deterministic Url and HashTag Segmentation
Segmenting Domain Names and Twitter HashTags into English language words.
Solve Challenge
NEED HELP?
View discussions
View top submissions
Submitted Code
Language: Python 3
Open in editor
import difflib
def correct_query(query, word_list):
words = query.split()
corrected_words = []
for word in words:
if word in word_list:
corrected_words.append(word)
else:
closest_match = difflib.get_close_matches(word, word_list, n=1)
if closest_match:
corrected_words.append(closest_match[0])
else:
corrected_words.append(word)
return ' '.join(corrected_words)

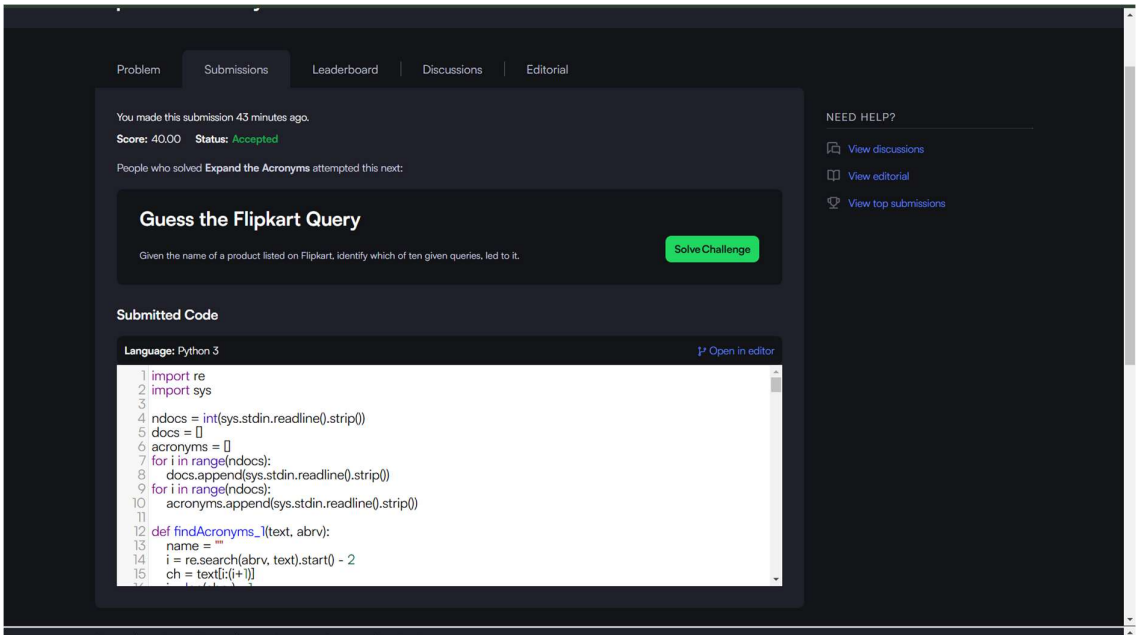

HackerRank
Prepare > Artificial Intelligence > Natural Language Processing > Deterministic Url and HashTag Segmentation
Exit Full Screen View
Problem
Submissions
Leaderboard
Discussions
Deterministic Url and HashTag Segmentation
This problem will introduce you to the segmentation of Domain Names and Social Media HashTags, into English Language words. To give you a quick idea of what segmentation means, here are a few examples of Domain names and Hash Tags which have been segmented.
Domain Name Examples:-
www.checkdomain.com => [check domain]
www.bigrock.com => [big rock]
www.namecheap.com => [name cheap]
www.appledomains.in => [apple domains]
Twitter Hash Tag Examples:
#honestyhour => [honesty hour]
#beinghuman => [being human]
#followback => [follow back]
#socialmedia => [social media]
#30secondstoearth => [30 seconds to earth]
The segmentation should be based on the list of 5000 most common words from here
Apart from the words in this list, you should also pick up numbers (both integer and decimal) like 100, 200.10 etc.
At this stage, we are going to use a very simple algorithm for the process. In case the input is a domain name, ignore the www. and/or the extensions (.com,.edu,.org,.in, etc.) In case the input is a hashtag, ignore the first # symbol. Split the input string, into a sequence of tokens. A token can either be:
A word in from the provided lexicon/dictionary.
An integer or decimal number.
Change Theme
Language
Python 3
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
import re
def split_words(line, tokens, regexps):
#print('line', line, 'tokens', tokens)
if not line:
return tokens
else:
for regexp in regexps:
m = regexp.match(line)
if m:
matched = m.group(0)
suffix = line[len(matched):]
new_tokens = tokens + [matched]
ans = split_words(suffix, new_tokens, regexps)
if ans:
return ans
return None
def main():
with open('words.txt') as f:
regexps = [re.compile(r'\d+(?:\.\d+)?')]
for w in sorted(re.split(r'[\n]+', f.read()), key=len, reverse=True):
if w:
regexps.append(re.compile(w, flags=re.IGNORECASE))
Line: 50 Col: 1
Upload Code as File
Test against custom input
Run Code
Submit Code


Prepare > Artificial Intelligence > Natural Language Processing > Disambiguation: Mouse vs Mouse
Disambiguation: Mouse vs Mouse
Points: 278.62 Rank: 3339
Problem
Submissions
Leaderboard
Discussions
You made this submission 3 days ago.
Score: 30.34 Status: Accepted
People who solved Disambiguation: Mouse vs Mouse attempted this next:
Language Detection
Detect the language for a given snippet of text.
Solve Challenge
NEED HELP?
View discussions
View top submissions
Submitted Code
Language: Python 3
Open in editor
def classify_mouse_context(sentence):
computer_keywords = ["input device", "click", "scroll", "computer", "USB", "wireless", "keyboard", "monitor"]
animal_keywords = ["tail", "fur", "cheese", "rodent", "genome", "postnatal", "development"]
sentence_lower = sentence.lower()
for keyword in computer_keywords:
if keyword in sentence_lower:
return "computer-mouse"
for keyword in animal_keywords:
if keyword in sentence_lower:
return "animal"
https://www.hackerrank.com/challenges/language-detection


HackerRank
Prepare
Certify
Compete
Apply
Search
Prepare > Artificial Intelligence > Natural Language Processing > Language Detection
Language Detection
Points: 278.62 Rank: 3339
Problem
Submissions
Leaderboard
Discussions
You made this submission an hour ago.
Score: 23.44 Status: Wrong Answer
NEED HELP?
View discussions
View top submissions
Submitted Code
Language: Java 8
Open in editor
import java.io.*;
import java.util.*;
public class Solution {
public static void main(String[] args) {
Scanner scanner = new Scanner(System.in);
StringBuffer buffer = new StringBuffer();
while (scanner.hasNextLine()) {
buffer.append(scanner.nextLine());
}
System.out.println(getGuessedLanguage(buffer.toString()));


Prepare > Artificial Intelligence > Natural Language Processing > The Missing Apostrophes
The Missing Apostrophes
Points: 278.62 Rank: 3339
Problem
Submissions
Leaderboard
Discussions
You made this submission an hour ago.
Score: 34.84 Status: Accepted
People who solved The Missing Apostrophes attempted this next:
Segment the Twitter Hashtags
Given a Twitter hashtag, segment it into words.
Solve Challenge
NEED HELP?
View discussions
View top submissions
Submitted Code
Language: Python 3
Open in editor
import zlib, base64, re
def getWordsEncrypted():
text = open('words.txt').read().lower().replace('\n', ' ')
text += open('book1.txt').read().lower().replace('\n', ' ')
text += open('book2.txt').read().lower().replace('\n', ' ')
text += open('book3.txt').read().lower().replace('\n', ' ')
text += open('book4.txt').read().lower().replace('\n', ' ')
text += open('book5.txt').read().lower().replace('\n', ' ')
text += open('book6.txt').read().lower().replace('\n', ' ')
text += " killers wearables scenarios repellents chemicals we've sensors hormones bursts trails cells markers cancers alerts they'll it's proteins indicators parkinson's lacks airports pei's beginnings vaccines tests impacts fans researchers malarial recognises "
text = text.replace("is'", "is ").replace('were', '').replace("parkinsons", "")
Snipping Tool
Screenshot copied to clipboard
Automatically saved to screenshots folder.
Mark-up and share


Prepare > Artificial Intelligence > Natural Language Processing > Segment the Twitter Hashtags
Segment the Twitter Hashtags
Points: 278.62 Rank: 3339
Problem
Submissions
Leaderboard
Discussions
You made this submission 39 minutes ago.
Status: Processing
NEED HELP?
View discussions
View top submissions
Submitted Code
Language: Python 3
Open in editor



Problem | Submissions | Leaderboard | Discussions | Editorial

You made this submission 43 minutes ago.

Score: 40.00 **Status:** Accepted

People who solved **Expand the Acronyms** attempted this next:

Guess the Flipkart Query

Given the name of a product listed on Flipkart, identify which of ten given queries, led to it.

Solve Challenge

NEED HELP?

View discussions

View editorial

View top submissions

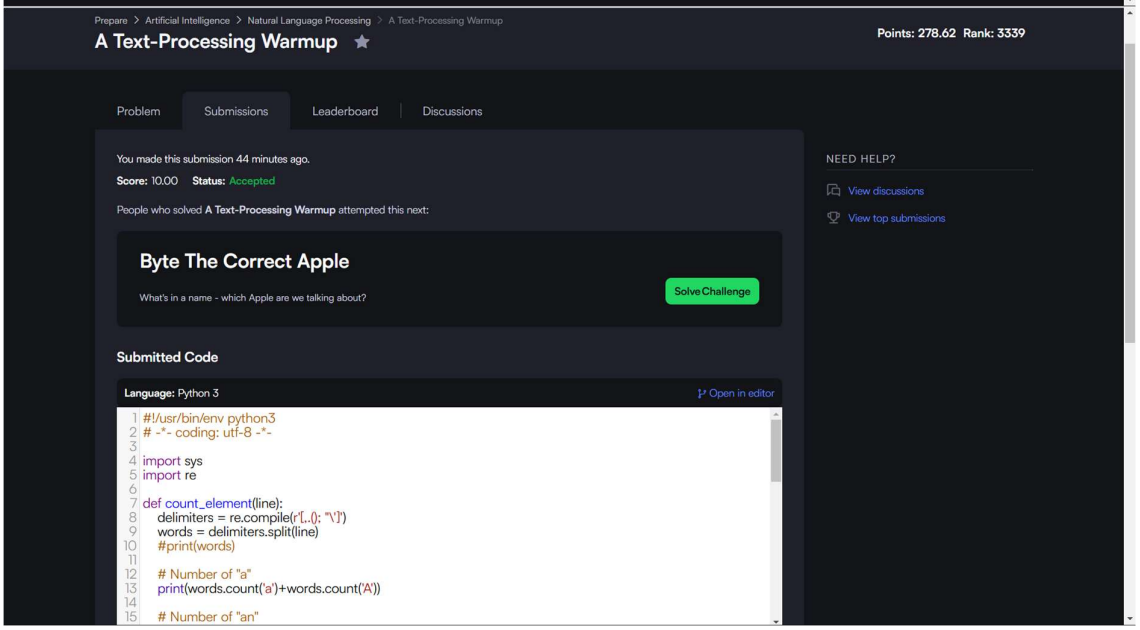
Submitted Code

Language: Python 3 — Open in editor

```
import re
import sys

ndocs = int(sys.stdin.readline().strip())
docs = []
acronyms = []
for i in range(ndocs):
    docs.append(sys.stdin.readline().strip())
for i in range(ndocs):
    acronyms.append(sys.stdin.readline().strip())

def findAcronyms_1(text, abrv):
    name = ""
    i = re.search(abrv, text).start() - 2
    ch = text[i:(i+1)]
```



Prepare > Artificial Intelligence > Natural Language Processing > A Text-Processing Warmup

A Text-Processing Warmup

Points: 278.62 Rank: 3339

Problem | Submissions | Leaderboard | Discussions

You made this submission 44 minutes ago.

Score: 10.00 **Status:** Accepted

People who solved **A Text-Processing Warmup** attempted this next:

Byte The Correct Apple

What's in a name - which Apple are we talking about?

Solve Challenge

NEED HELP?

View discussions

View top submissions

Submitted Code

Language: Python 3 — Open in editor

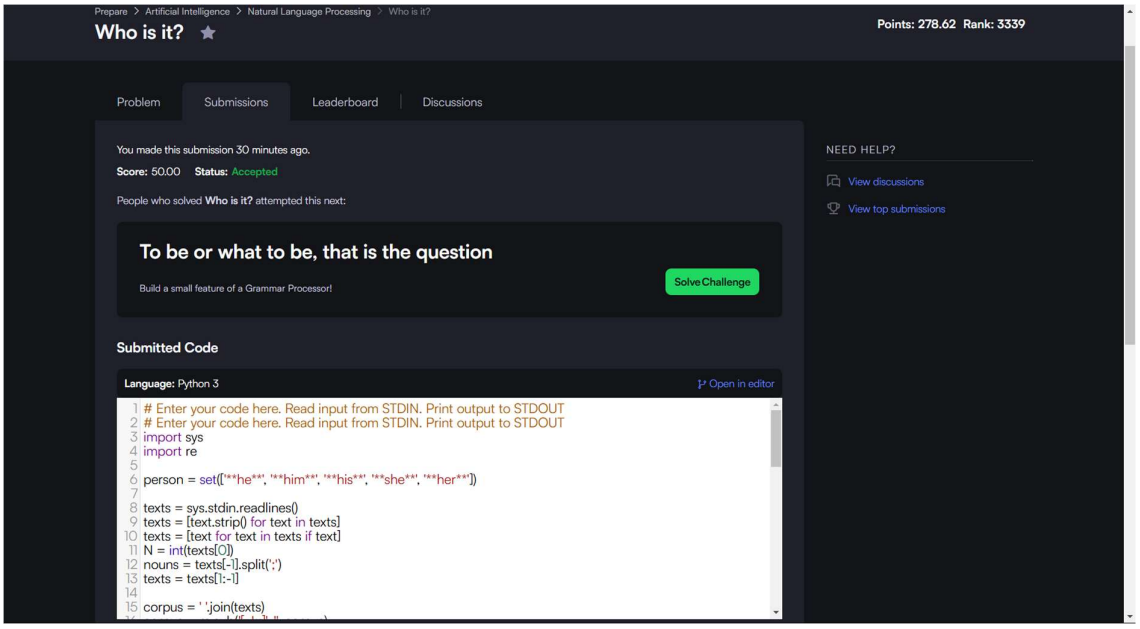
```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-

import sys
import re

def count_element(line):
    delimiters = re.compile(r'[.,(); "\']')
    words = delimiters.split(line)
    #print(words)

    # Number of "a"
    print(words.count('a')+words.count('A'))

    # Number of "an"
```


Prepare > Artificial Intelligence > Natural Language Processing > Who is it?
Who is it?
Points: 278.62 Rank: 3339
Problem
Submissions
Leaderboard
Discussions
You made this submission 30 minutes ago.
Score: 50.00 Status: Accepted
People who solved Who is it? attempted this next:
To be or what to be, that is the question
Build a small feature of a Grammar Processor!
Solve Challenge
NEED HELP?
View discussions
View top submissions
Submitted Code
Language: Python 3
Open in editor
# Enter your code here. Read input from STDIN. Print output to STDOUT
# Enter your code here. Read input from STDIN. Print output to STDOUT
import sys
import re
person = set(['**he**', '**him**', '**his**', '**she**', '**her**'])
texts = sys.stdin.readlines()
texts = [text.strip() for text in texts]
texts = [text for text in texts if text]
N = int(texts[0])
nouns = texts[-1].split(';')
texts = texts[1:-1]
corpus = ' '.join(texts)