

Extracció paraules fitxer text

20 de septiembre de 2024

```
import io
import re
import time

patro= re.compile("(?!\\S)[a-zA-ZÀÈÌÒÙÁÉÍÓÚÏÜàèìòùáéíóúçÇïüñÑ\\.]+(?!\\S)")

fitxer = io.open("elpuntavui.txt","r", encoding ="utf-8")

llista_linies = fitxer.readlines()

print(llista_linies)

# llista_linies = [linia1, linia2, linia3,.....]

for linia in llista_linies:
    print(linia)
    print(patro.findall(linia))
    time.sleep(3)
```



It seems like you're attempting to read a text file, extract words that match a certain regular expression, and then print both the lines and the words that match.

Let me break down your code a little:

- **patro** : This is the compiled regular expression, where you're looking for words that contain alphabetic characters, including accented characters. The expression uses negative lookbehind and negative lookahead to ensure that the match is not preceded or followed by non-whitespace characters (essentially matching whole

Comienza con ChatGPT