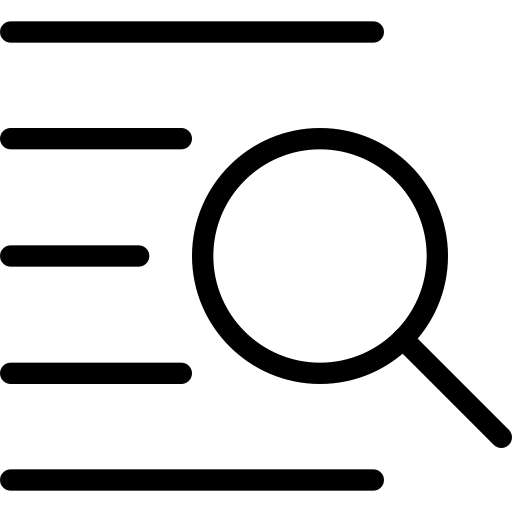


Γλωσσική

Τεχνολογία

**Project**

2021 - 2022



**Στυλιανός Στυλιανάκης**  
**1059713**

[**Link to code**](https://github.com/Steliostyl/Language-Technology-Project-2022)

## ΜΕΡΟΣ Α

### Προσκομιστής Ιστοσελίδων

Mε τη χρήση του **Scrapy** δημιουργήθηκε ο Crawler, με τη βοήθεια του οποίου ανακτώνται άρθρα από τις ειδησεογραφικές ιστοσελίδες *CNN* και *CNBC.* Πιο συγκεκριμένα, μέσα στο αρχείο **spiders.py** ορίστηκαν δύο **spiders**, το **CnbcSpider** και το **CnnSpider,** τα οποία είναι υπεύθυνα για την εξαγωγή άρθρων από τις αντίστοιχες ιστοσελίδες.