

Laborator 10: Prelucrarea limbajului natural

Utilizați Wordnet pentru a înlocui toate substantivele dintr-un text cu hypernym-ul direct, pentru primul sens găsit. Calculați distanța semantică maximă între cuvintele din fiecare propoziție.

Cerințe:

(0.2) Descărcați [Wordnet](#). Identificați un API în limbajul de programare preferat. Configurați API-ul pentru resursa locală Wordnet descărcată anterior.

(0.2) Folosiți acel API pentru a identifica primul hypernym al unui termen.

(0.3) Implementați un program care primește la intrare calea către un fișier *.txt (conținutul este text în limba engleză), parcurge conținutul fișierului și caută fiecare cuvânt în Wordnet. Când găsește un cuvânt, verifică dacă una din părțile de vorbire posibile este substantiv. Dacă da, caută primul synset hypernym al primului sens pentru acel substantiv. Primul cuvânt din acel synset va fi scris în fișierul de ieșire în locul cuvântului original. Ieșirea va fi scrisă în același fișier din care ați citit textul inițial.

(0.3) Pentru fiecare pereche de cuvinte din fiecare propoziție calculați distanța semantică dintre primele synset-uri găsite pentru cele două cuvinte. În ieșire, după fiecare propoziție scrieți distanța maximă determinată între cuvintele din acea propoziție.

Exemplificare (pentru Python - NLTK):

```
from nltk.corpus import wordnet
syns = wordnet.synsets("dog")
print(syns)
```

afișează

```
[Synset('dog.n.01'), Synset('frump.n.01'), Synset('dog.n.03'), Synset('cad.n.01'),
Synset('frank.n.02'), Synset('paw1.n.01'), Synset('andiron.n.01'), Synset('chase.v.01')]
```

Primul synset din listă are eticheta "n", deci este substantiv. Primul cuvânt din primul hypernym poate fi obținut cu `syns[0].hypernyms()[0].lemmas()[0]`.

Pentru determinarea distanței semantice folosiți `shortest_path_distance` de aici: https://www.nltk.org/_modules/nltk/corpus/reader/wordnet.html (documentația pentru ultima versiune e aici: https://www.nltk.org/_modules/nltk/corpus/reader/wordnet.html).