Jegyzőkönyv

IET HF1: Szemantikus keresés

Bendicsek Márton Bendegúz

Github: https://github.com/bendicsekb/semantic\_web.git

Feladat 1 indexelés

Egyszerű indexelést pythonban dictionarykkel érdemes csinálni, hiszen az már egy indexelt adatstruktúra. A fájlokat beolvasva szavakra és fájlokra is raktam indexet.

A program használatához a create\_index.py-t a tartalmazó mappából kell futtatni (python 3.8), két kikötéssel:

az indexelendő szöveg kiterjesztése nem lehet .json

a fájlok a create\_index.py mappájától relatív “../res/indexing” mappában legyenek

A program felépít egy indexet, majd elmenti a simple\_indexes.json fájlba.

Példa az indexre (python annotációkkal):

*{*

*'words': { 'john': {'document1': 3, 'document2': 5},*

*'doe' : {'document1': 1, 'document3': 2}},  
'documents': { 'document1': {'foo':2, 'bar':1 },*

*'document2':{'asd': 1, 'bsd': 5}}*

*}*

Feladat 2 kezelés

A második feladathoz írt program a már meglévő indexet beolvassa, az indításkor paraméterként kapott szavak dokumentumbeli gyakorisága alapján csökkenő sorrendben kiírja azokat a dokumentumokat melyekben minden keresőszó szerepel. Az összes szó dokumentumbeli előfordulását szummáztam dokumentumonként.

Példa: python my\_search.py Sony utolsó napjaira

Kimenet:

Sony utolsó napjaira

34914.txt

A feladat során felhasználtam a dokumentum és a szó alapú indexelést is.

Feladat 3 szemantikus keresés