

Versiunea paralela

-Threads-

~Proiect APD~

Safta Maria-Andreea

### Tema proiect : Text Processing

Pentru realizarea versiunii paralele a acestei tema m-am folosit de functia de tokenizare , normalizare a unui text si de metoda cu fire multiple de executie , threads , iar ca limbaj de programare am folosit C++.

Programul citeste dintr-un fisier text , input.txt , continutul acestuia , care cu ajutorul functiei “normalizeText” transforma toate caracterele in litere mici pentru a nu se face diferente de capitalizare . Functia “tokenize” prezinta un flux de string “std::istringstream” , cu ajutorul caruia citeste textul cuvant cu cuvant si apoi elimina caracterele de punctuatie “std::remove\_if”.Aceasta functie se foloseste de regula pentru o mai buna analiza a cuvintelor .

Functia “countWordFrequenciesThread” primeste un text ca si argument si contorizeaza frecventa fiecarui cuvant utilizand fire multiple de executie . “std::unordered\_map<std::string, int> wordFreq;” reprezinta o harta neordonata in care se stocheaza cuvintele si frecventele lor , apoi textul este convertit in litere mici , eliminand si semnele de punctuatie . Pentru protejarea accesului la harta de frecvente cand aceasta se incearca a fi modificata de mai multe fire de executie se foloseste Mutex , “std::mutex mtx;”. Functia lambda este folosita pentru a contoriza frecventa cuvintelor intr-un interval . Pentru inceput se parcurg token-urile din intervalul specificat , apoi se incrementeaza contorul pentru fiecare cuvant gasit si apoi se actualizeaza harta locala . Se calculeaza numarul aproximativ de cuvinte generate pentru fiecare fir de executie apoi se stocheaza intr-un vector . Vectorul contine perechi de cuvinte si frecventele lor din harta globala . In final se deschide fisierul txt , continutul se citeste se afiseaza rezultatele si apoi se afiseaza si timpul de executie in milisecunde .

Date experimentale :

Frecventa cuvintelor este:

```
a: 16
abordare: 1
abstracta: 1
accent: 1
aceasta: 2
acestui: 1
actori: 1
adanci: 1
adesea: 2
aducand: 1
aduce: 1
adus: 2
ai: 1
al: 5
ale: 1
antrenamentele: 1
antreneaza: 1
are: 3
armonios: 1
arta: 6
artei: 1
```

```
vibranta: 1
vie: 1
viziune: 1
vorba: 1
william: 1
xixlea: 1
xx: 1
xxi: 1
Timpul de executie: 71.9743 milisecunde
```

C:\Users\Andreea\Desktop\AN3\_sem2\APD\TextProc

```
viziune: 1
vorba: 1
william: 1
xixlea: 1
xx: 1
xxi: 1
Timpul de executie: 77.1859 milisecunde
```

C:\Users\Andreea\Desktop\AN3\_sem2\APD\TextProc