

Tema 2 SPRC MPI

		mappers		
		2	3	4
reducers	4	7.3	5.29	4.8
	6	5.2	4.36	4.32
	8	4.4	4.17	4.39
	10	4.1	4.1	5

TEMA SE POATE RULA FOLOSIND MAKEFILEUL

Detalii de implementare:

Programul principal imparte fisierul in mod egal la mapperi.

Acestia primesc offsetul de la care trebuie sa citeasca fisierul si lungimea stringului aferent.

Fiecare mapper (cu exceptia ultimului) citesc un string de o lungime un pic mai mare decat cea primita (pentru a rezolva cazul cand offsetul primit este chiar in mijlocul unui cuvant)

Astfel, fiecare mapper ajusteaza prefixul si sufixul stringului citit:

se elimina separatorii si se intregeste ultimul cuvant.

In continuare fiecare mapper creeaza reducerii aferenti si imparte echitabil stringul. Si de aceasta data maperul are grija sa nu trimita reducerilor franturi de cuvant.

Reducerii primesc un string si il parseaza (se folosesc expresii regulate si sscanf).

Structura de date folosita pentru calcularea frecventelor este hashmap

(map in C++, care foloseste ca si cheie stringul)

Serializarea hashmap-ului pentru a trimite rezultatul la mapper se face printr-un vector de structuri (se creeaza un MPI_Type pentru structura)

La fel procedeaza si mapperii, care dupa ce centralizeaza rezultatele de la reduceri trimit rezultatul la programul principal.

Programul principal centralizeaza si sorteaza rezultatele in functie de frecventa si scrie in fisier.

Observatii:

Timpii mari de executie se cauzeaza manevrarii stringurilor (care dureaza mult mai mult decat daca s-ar fi lucrat direct cauza trecerii cu datele din fisier) si din datelor dintr-o structura in alta (din hash map in structura, din structura in vector)