

Sistemes Operatius II

Pràctica 3

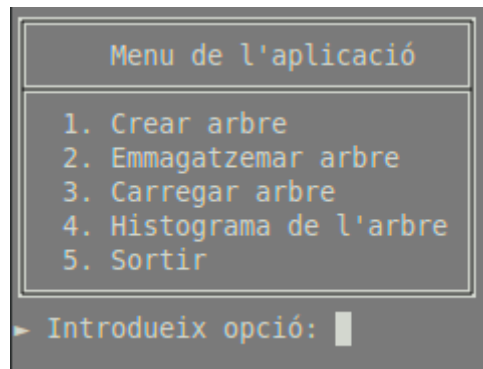
Vicent Roig / Igor Dzinka

17/11/2014

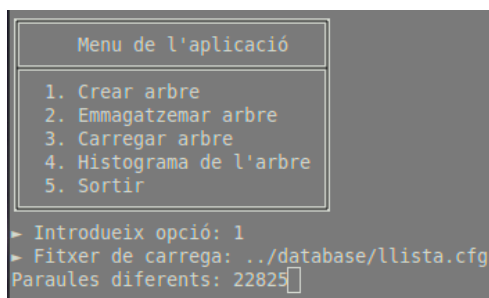
1. El menú de l'aplicació

Es demana desenvolupar una interfície de menú senzilla que permeti que l'usuari interactui amb l'aplicació.

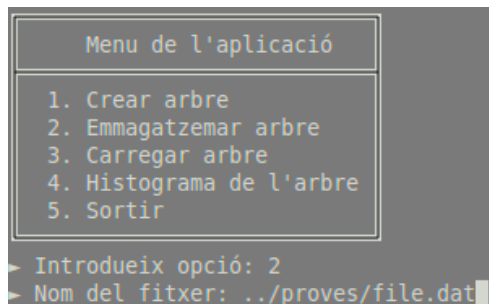
Hem creat un menú amb 5 opcions que es demanen a l'enunciat de la pràctica:



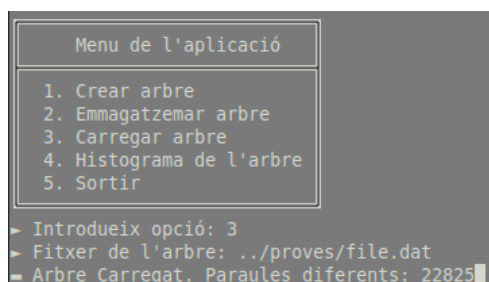
IMPORTANT: Quan s'executa una opció del menú un cop finalitzada l'execució, per tornar al menú principal cal prémer ENTER.



Crear arbre - opció que permet crear un arbre a partir d'un fitxer de base de dades. Un cop seleccionada aquesta opció, l'aplicació demana que l'usuari entri el nom del fitxer de la base de dades. Si el nom del fitxer es correcte, l'aplicació procedeix a extreure dades de la base de dades i les guarda dins d'un RBTtree. Un cop finalitzat el processament, s'escriu per pantalla el missatge d'èxit i el nombre de paraules úniques extretes de la base de dades.



Emmagatzemar arbre – opció que permet guardar l'arbre al disc. Quan es selecciona aquesta opció i existeix un arbre que es pugui l'aplicació demana a l'usuari el nom del fitxer al que vol guardar l'arbre. Un cop fet això l'aplicació realitza l'escriptura de l'arbre al fitxer.



Carregar arbre – permet recuperar l'arbre d'un fitxer. Quan es selecciona l'opció, el sistema demana a l'usuari el nom del fitxer del que vol recuperar les dades. Si el nom del fitxer es correcte, si encara existeix un arbre creat, es buida. L'aplicació procedeix a llegir les dades i les guarda dins d'un RBTtree. Un cop finalitzat el processament, s'escriu per pantalla el missatge d'èxit i el nombre de paraules úniques dins l'arbre extret.

```
Menu de l'aplicació
1. Crear arbre
2. Emmagatzemar arbre
3. Carregar arbre
4. Histograma de l'arbre
5. Sortir
> Introdueix opció: 4
■ Grafica 'treeHistogram.svg' generada.
```

Histograma de l'arbre – permet recopilar dades estadístiques i representar-les amb GNUPLOT, per tal d'obtenir una gràfica de l'histograma de la longitud de les paraules que conté l'arbre.

```
Menu de l'aplicació
1. Crear arbre
2. Emmagatzemar arbre
3. Carregar arbre
4. Histograma de l'arbre
5. Sortir
> Introdueix opció: 5
Exit
```

Sortir – finalitza l'aplicació.

La persistència de l'arbre.

Per tal d'evitar haver de crear l'arbre cada cop que s'inicialitza l'aplicació, es proposa poder emmagatzemar i carregar l'estructura d'arbre a/de disc.

Inicialment vàrem plantejar la forma de desar l'arbre mitjançant escriptura de paraules de mida fixa MAX_WORDCHR, això ens generava problemes i en tot cas el fet de tenir aquests problemes ens va portar a rellegir el pdf i trobar un plantejament correcte i molt més òptim:

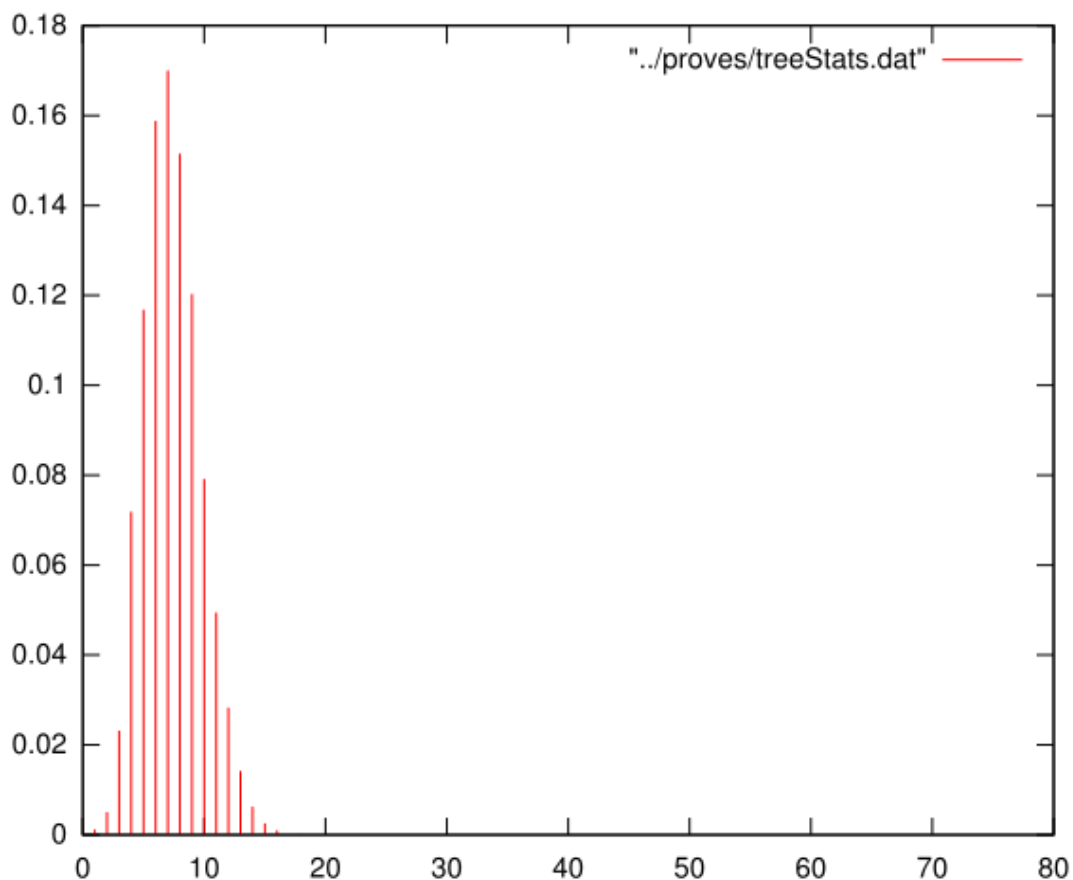
En lloc d'escriure una paraula de mida fixa cada cop, escriurem únicament els bytes necessaris, amb l'única variació de que amb aquest mètode, a més tenim que tenir en compte el fet de guardar per a cada paraula la seva longitud (si no ho fem així no seríem capaços de carregar-ho posteriorment).

Com hem pogut observar, les paraules que apareixen més freqüentment solen ser les de 8-10 caràcters, per tant, el fet de haver guardat per cada paraula una mida fixa de 75 caràcters deixa en evidència que no era una forma òptima d'afrontar el problema. De fet, encara que amb aquest plantejament hem de guardar una nova dada (*l'enter amb la longitud de la paraula, típicament 4 bytes*), segueix essent en promig una solució molt millor.

La comunicació interprocés mitjançant canonades.

L'objectiu és analitzar la longitud de les paraules que hi ha a l'arbre i realitzar una gràfica mitjançant GNUPLOT.

A continuació mostrem l'histograma obtingut d'un subconjunt de 10 fitxers de la base de dades. Per tal de que es pugui fer una comprovació, l'adjuntem al lliurament com directori data base. Tots els resultats del valgrind.txt i altres fitxers del directori provenen d'aquest subconjunt.



Un dels problemes al que ens hem enfrontat a aquesta part de la pràctica es que fins ara, en cap moment havíem guardat la longitud de la paraula més llarga que hi ha a l'arbre. El fet de no tenir aquesta dada no ens permetia a priori definir la mida de l'array que utilitzarem per generar l'histograma.

Posteriorment ens hem adonat que no era necessari disposar de la mida màxima de la paraula, donat que en tot cas, aquesta tindrà exactament la mida màxima definida en pràctiques anteriors (MAX_WORDCHR, en nostre cas 75 caràcters), d'aquesta manera no cal tenir en compte si efectivament la mida màxima es aquesta o inferior, sinó que tindrem un histograma amb un marge de confiança.

Per tant, als resultats obtinguts sabem que hi haurà una certa quantitat de posicions reservades que estaran, molt probablement, per sobre de la longitud de paraula més llarga realment trobada (recordem que a la pràctica 2 vàrem obtenir com a paraula més llarga de la base de dades una amb mida 69).