

# Sistemes Operatius II - Pràctica 3

Octubre del 2017

La tercera pràctica se centra en en la persistència de l'estructura de l'arbre, és a dir, en poder desar i carregar la informació emmagatzemada a/de disc.

## Índex

<b>1</b>	<b>La pràctica</b>	<b>2</b>
1.1	Interfície de menú . . . . .	2
1.2	Creació de l'arbre . . . . .	2
1.3	Emmagatzemament de l'arbre a disc . . . . .	3
1.4	Consulta d'informació emmagatzemada a l'arbre . . . . .	4
<b>2</b>	<b>Implementació</b>	<b>4</b>
<b>3</b>	<b>Entrega</b>	<b>4</b>

# 1 La pràctica

Aquesta pràctica se centra en

- Utilitzar una interfície de menú senzilla que permeti que l'usuari interactuï amb l'aplicació, veure secció 1.1. En aquesta pràctica es proporciona a l'estudiant del codi C d'aquesta interfície.
- Permetre la creació de l'arbre a partir de múltiples fitxers PDF, veure secció 1.2.
- 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 quan l'usuari vulgui, veure secció 1.3.
- Permetre fer consultes respecte la informació emmagatzemada a l'arbre, veure secció 1.4.

Es descriuen a continuació cadascun dels anteriors punts.

## 1.1 Interfície de menú

Per tal de facilitar la interacció de l'usuari amb l'aplicació es proposa desenvolupar una interfície de menú textual. El menú ha d'incloure les següents 5 opcions:

1. Creació de l'arbre. La creació de l'arbre correspon a la pràctica 2. En cas que hi hagi un arbre prèviament creat caldrà alliberar la memòria associada. Veure secció 1.2
2. Emmagatzemament de l'arbre. En seleccionar l'usuari aquesta opció l'aplicació haurà de demanar per teclat el nom del fitxer on es desarà l'arbre. A continuació l'aplicació desarà en aquest fitxer la informació dels nodes l'arbre mitjançant funcions d'entrada/sortida no formatada. A la secció 1.3 s'especifica el format de fitxer en què s'emmagatzema l'arbre.
3. Lectura de l'arbre. En seleccionar l'usuari aquesta opció l'aplicació haurà de demanar per teclat el nom del fitxer amb l'arbre a llegir. L'aplicació llegirà del fitxer fent i crearà l'estructura d'arbre, veure secció 1.3. En cas que hi hagi un arbre prèviament creat caldrà alliberar la memòria associada.
4. Permetre fer consultes respecte la informació emmagatzemada a l'arbre, veure secció 1.4.
5. Sortida. Mitjançant aquesta opció s'allibera tota la memòria reservada i se surt de l'aplicació.

## 1.2 Creació de l'arbre

En seleccionar l'usuari aquesta opció l'aplicació haurà de demanar per teclat el fitxer de text pla que conté un llistat dels fitxers PDF a analitzar. L'aplicació crearà l'arbre a partir de tots els PDFs. En finalitzar tornarà a mostrar el menú amb totes les opcions.

El format de fitxer és similar al que s'ha fet servir a la pràctica 1. En particular, a la primera línia s'especifica el nombre de fitxers PDF que s'especifiquen a continuació. Després es llisten els fitxers PDFs, un a cada línia. S'hauran d'analitzar cadascun dels fitxers PDF introduïnt la informació en un únic arbre.

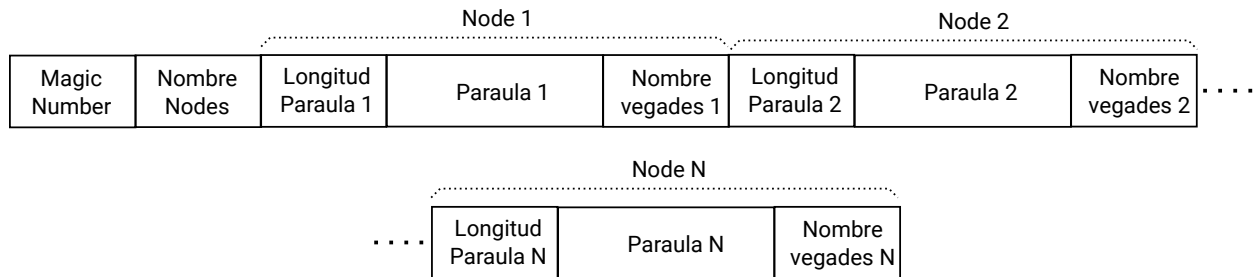


Figura 1: Format del fitxer per emmagatzemar la informació emmagatzemada a l'arbre.

### 1.3 Emmagatzemament de l'arbre a disc

L'emmagatzemament de l'arbre a disc evita haver de crear l'arbre cada cop que s'engegui l'aplicació. Mitjançant el menú es podrà decidir si es vol desar o carregar la informació de l'arbre a disc. La informació s'haurà de desar mitjançant la tècnica no formatada, fent servir el format de fitxer especificat a la figura 1. S'especifica el format de fitxer per assegurar interoperabilitat entre els formats generats per totes les aplicacions dels vostres companys.

Aquest és el detall de l'estructura del fitxer

- Nombre màgic, 4 bytes. El nombre màgic del fitxer és 0x0133C8F9<sup>1</sup>.
- Nombre de nodes  $N$  de l'arbre, 4 bytes, un sencer amb signe.
- Per a cada node s'emmagatzema la següent informació: a) nombre de bytes de la paraula, 4 bytes amb signe, b) la paraula (sense el byte 0 final de la cadena), c) el nombre de vegades que la paraula ha aparegut en analitzar els fitxers PDF, 4 bytes amb signe. Els nodes es poden emmagatzemar a disc en l'ordre que es vulgui (i.e. es pot començar pel node arrel o bé pels nodes fills).

No s'han d'invertir els bytes per assegurar compatibilitat entre arquitectures big i little-endian. Observar que al fitxer només es desa la informació emmagatzemada a cada node. No s'hi desa informació associada a l'estructura de l'arbre (pares, fills o germans de cada node).

En llegir la informació de l'arbre es llegirà el fitxer amb el format especificat. Atès que només la informació de cada node és possible que l'estructura de l'arbre resultant sigui diferent de l'estructura que hi havia en el moment de desar-lo. Això no importa pas ja que l'estructura de l'arbre no té importància per aquestes pràctiques.

La lectura de fitxer ha de ser robusta en front a possibles errors. En cas que es produeixi algun dels errors esmentats a continuació, cal mostrar un missatge per pantalla, alliberar tota la memòria que faci falta, i tornar a mostrar el menú. En cas que es llegeixi un nombre màgic no vàlid, s'aborta la lectura del fitxer. En cas que es llegeixin nombres sencers negatius (o igual a zero), es pot abortar la lectura del fitxer (tornant a mostrar el menú) o saltar-se la informació del node llegit i continuar amb el següent node. Podeu suposar que el fitxer està complet (i.e. hi ha tots els nodes especificats al camp de nodes de l'arbre).

<sup>1</sup>Aquest nombre surt d'agafar la data 2017/10/01 i "transformar" el nombre 20171001 a hexadecimal.

## 1.4 Consulta d'informació emmagatzemada a l'arbre

L'usuari haurà de poder introduir per teclat una paraula qualsevol. L'aplicació buscarà aquesta paraula a l'arbre i imprimirà el nombre de vegades que la paraula ha aparegut en analitzar els fitxers PDF. Aquesta funcionalitat s'introdueix per comprovar que en llegir l'arbre de disc es llegeixen les dades de forma correcta.

## 2 Implementació

Aquesta secció dona alguns consells per tal d'implementar correctament aquesta pràctica.

Es proporciona a l'estudiant del codi en C del menú a utilitzar. No s'ha inclòs el codi associat a cadascuna de les opcions possibles. Tingueu en compte que l'usuari pot escollir les opcions del menú en qualsevol ordre i, per tant, caldrà tenir en compte les restriccions associades. Per exemple, a l'hora de desar l'arbre cal assegurar que hi ha un arbre a memòria; a l'hora de llegir un arbre de disc cal alliberar l'arbre en cas que n'hi hagi algun carregat a memòria, etc.

1. Implementeu la funció que permet crear l'arbre. En aquesta pràctica la creació de l'arbre es fa a partir de múltiples fitxers PDF, vegeu 1.2. Se suposa que el fitxer de text pla que conté el llistat de fitxers PDF està correctament formatat: hi ha tants fitxers PDF com s'indiquen a la primera línia del fitxer. S'ha adjuntat amb l'enunciat un conjunt de fitxers PDF que es poden fer servir per fer proves. Copieu aquests fitxers dintre del directori que conté el vostre codi.
2. Implementeu la funció que permet consultar el nombre de vegades que apareix una paraula a l'arbre, veure secció 1.4. Si la paraula no es troba a l'arbre, mostreu un missatge d'error per pantalla.
3. Implementació de les funcions per desar i carregar la informació de l'arbre a disc. L'arbre es desarà en un únic fitxer tal com s'ha comentat a la secció 1.3. Per a desar l'arbre es recomana fer servir la instrucció *fwrite* i per llegir la instrucció *fread*, veure fitxa del campus. Us farà falta recòrrer tot l'arbre: baseu-vos en la funció que esborra l'arbre per implementar-ho. Se us proporciona un fitxer `arbre.so2` perquè pugueu fer proves en llegir el fitxer en el format especificat: la paraula “younger” apareix 82 vegades, i la paraula “turtle” 64 vegades.

Un cop hagueu implementat l'escriptura i lectura de l'arbre de disc, assegureu-vos que funciona correctament fent servir el *valgrind*.

## 3 Entrega

El fitxer que entregueu s'ha d'anomenar `P3_NomCognom1NomCognom2.tar.gz` (o `.zip`, o `.rar`, etc), on `NomCognom1` és el nom i cognom del primer component de la parella i `NomCognom2` és el nom i cognom del segon component de la parella de pràctiques. El fitxer pot estar comprimit amb qualsevol dels formats usuals (`tar.gz`, `zip`, `rar`, etc). Dintre d'aquest fitxer hi haurà d'haver tres carpetes: `src`, que contindrà el codi font, i `doc`, que contindrà la documentació addicional en PDF. Aquí hi ha els detalls per cada directori:

- La carpeta `src` contindrà el codi font. S'hi han d'incloure tots els fitxers necessaris per compilar i generar l'executable. El codi ha de compilar sota Linux amb la instrucció `make`. Editeu el fitxer *Makefile* en cas que necessiteu afegir fitxers C que s'hagin de compilar. El codi font ha d'estar comentat. No es necessari incloure els fitxers PDF amb el codi.
- El directori `doc` ha de contenir un document (tres o quatre pàgines, màxim cinc pàgines, en format PDF, sense incloure la portada) explicant el funcionament de l'aplicació, la discussió de les proves realitzades i els problemes obtinguts. En aquest document no s'han d'explicar en detall les funcions o variables utilitzades.

La data límit d'entrega està indicat al document de planificació. El codi té un pes d'un **80%** (codi amb funcions comentades, codi modular i net, ús correcte del llenguatge, bon estil de programació, el programa funciona correctament, tota la memòria és alliberada, sense accessos invàlids a memòria, etc.). Tingueu en compte que els professors comprovaran el bon funcionament del vostre codi fent servir el `valgrind` amb el fitxer `arbre.so2`, un fitxer que s'ha inclòs en aquesta pràctica. El document té un pes del **20%** restant.