



UNIVERSIDAD DE BUENOS AIRES

FACULTAD DE INGENIERIA

DEPARTAMENTO DE COMPUTACIÓN

75.06 – ORGANIZACIÓN DE DATOS

TRABAJO PRÁCTICO

BOOQUERIO

ETAPA 1

MARZO 2011

Índice General

1	Introducción	3
2	Enunciado	4
	Primera Etapa	5
2.1.1	<i>Parser</i>	5
2.1.2	<i>Organización de tipo Archivo de Registros Variables</i>	5
2.1.3	<i>Organización de tipo Archivo de Bloques</i>	5
2.1.4	<i>Árbol B+</i>	6
2.1.5	<i>Hash Extensible</i>	6
2.1.6	<i>Editorial</i>	6
2.1.7	<i>Autor</i>	6
2.1.8	<i>Título</i>	6
2.1.9	<i>Palabras</i>	6
3	Criterio de aprobación	7
3.1	Funcionalidad	7
3.2	Documentación	7
4	Referencias.....	9

1 Introducción

Este documento consiste en el enunciado del trabajo práctico de la asignatura. En el mismo se especifican los requerimientos de cada etapa de entrega, dejando de lado el cronograma de entregas que se encuentran en la página o grupo de correo de comunicación de la cátedra respectivamente.

Toda aclaración, indicación o respuesta a consultas (ofrecidas en clase o mediante el grupo yahoo) serán tomadas como extensión y parte explícita de este enunciado.

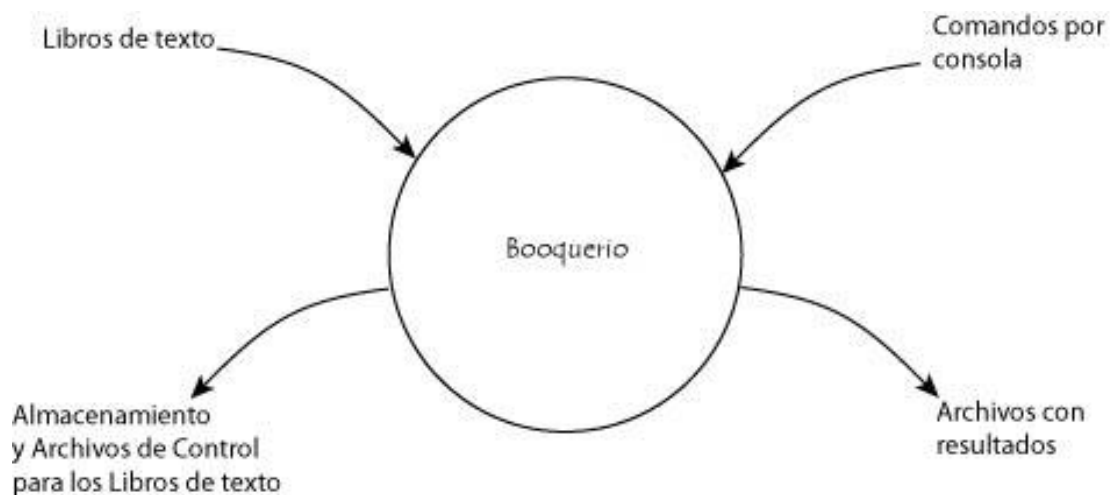
La forma de trabajo con los grupos es descripta en el Reglamento de Trabajos Prácticos de la Cátedra (http://materias.fi.uba.ar/7506C/blog/?page_id=9)

El presente enunciado describe el desarrollo de un sistema de Almacenamiento de Libros, para su consulta y lectura, llamado Booquerio.

2 Enunciado

El sistema se propone resolver el almacenamiento, consulta y obtención de libros de texto. Para lograr este objetivo se hará uso de los contenidos brindados en diferentes módulos de la materia.

La Primer Entrega se avocará a la aplicación de los conceptos de Organización de Archivos.



Libros de Texto = Son los Datos a almacenar y administrar. Se estructuran en formato texto plano, por ello cada grupo debe analizarlos para crear el interprete (parser) adecuado.

Almacenamiento y Archivos de Control para los Libros de Texto = Se encuentran todos dentro de un directorio, especificado a través de un archivo de configuración de la aplicación, y pueden tener jerarquía de subdirectorios interna. Es donde se guarda toda la información necesaria para poder funcionar.

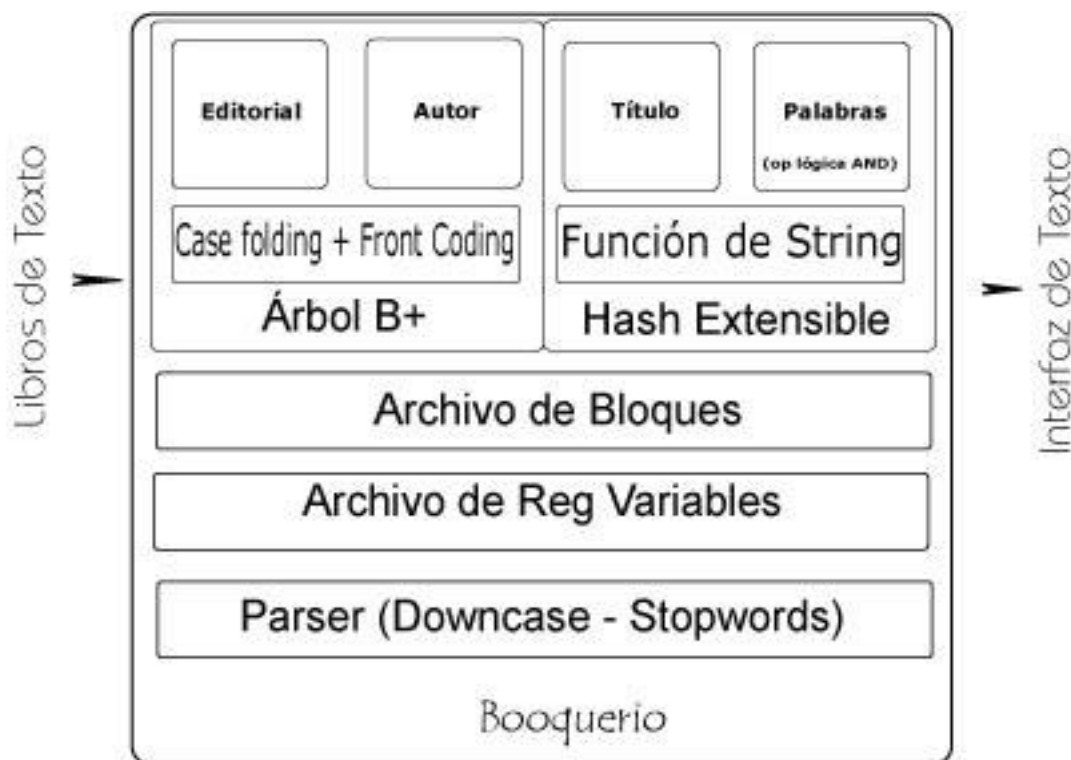
Comandos por consola = Toda la interacción del usuario con el sistema se realiza a través de comandos por consola.

Archivos con resultados = Como respuesta a toda interacción, el sistema generará archivos de respuesta en el directorio donde se llame a la aplicación.

La resolución del trabajo práctico debe ser realizada en plataforma Linux y lenguajes C o C++ (preferentemente respetando el estándar ANSI), y la entrega debe constar de un Makefile para su compilación. Además debe funcionar en calidad de Usuario (user) del sistema operativo.

Primera Etapa

El objetivo de esta etapa es resolver la aplicación, utilizando las herramientas brindadas por el módulo Organización de Archivos. Por ello se plantea la siguiente relación entre esos conceptos.



2.1.1 Parser

Se deberá implementar un módulo de interpretación o parser de los archivos de libros. Este mismo módulo identificará los datos contenidos y podrá ofrecer valores específicos (Editorial, Autor, Título, Palabras contenidas).

Pero en el caso de solicitar las palabras contenidas, se aplicarán dos técnicas de refinamiento, Downcase y Stopwords (tomando en cuenta los términos hallados en un archivo de configuración).

Por otras especificaciones recurran a los tutores.

2.1.2 Organización de tipo Archivo de Registros Variables

Se utilizarán primeramente como estructura para almacenar y asignar una identificación a cada libro procesado por el sistema.

Puede ser utilizada en otros casos.

2.1.3 Organización de tipo Archivo de Bloques

Se encargará de almacenar los árboles y tablas de dispersión (hash). El contenido de cada bloque o registro se deberá adecuar al tipo de dato a almacenar.

Puede ser utilizada en otros casos.

2.1.4 Árbol B+

Se implementará con las estructuras vistas en clases y se adaptará la clave para cada uso específico.

2.1.5 Hash Extensible

Se implementará según las estructuras vistas en clases y se adaptará el tipo de registro para cada uso específico.

2.1.6 Editorial

Es un índice de Clasificación de los Libros. De tipo B+.

2.1.7 Autor

Es un índice de Clasificación de los Libros. De tipo B+.

2.1.8 Título

Es un índice de identificación de los libros. De tipo Hash Extensible.

2.1.9 Palabras

Es un índice de selección de los libros. De tipo Hash Extensible.

3 Criterio de aprobación

Como se especifica en el Reglamento de la materia existe un criterio mínimo para poder acceder a una re-entrega en cada etapa. A continuación se menciona una lista de requerimientos que forman parte de dicho criterio. No cumplir con alguna de ellas implica no cumplir el mínimo requerido. Pero no vale la inversa, es decir, cumplir con ellas no implica cumplir con el criterio mínimo.

3.1 Funcionalidad

- *Archivo de configuración (con directorio de almacenamiento)*
- *Tomar Texto: ./ejecutable -i "archivo de texto"*
- *Procesar Editorial: ./ejecutable -e (procesa los no procesados)*
- *Procesar Autor: ./ejecutable -a*
- *Procesa Título: ./ejecutable -t*
- *Procesa Palabras: ./ejecutable -p*
- *Listar Archivos Tomados: ./ejecutable -l (muestra identificador, Título, Autor, Editorial y cantidad de palabras registradas para ese libro).*
- *Obtener Archivo: ./ejecutable -o ID_Archivo*
- *Quita Archivo: ./ejecutable -q ID_Archivo (se eliminan las entradas en los otros índices)*
- *Ver Estructura: Genera archivos en forma de texto plano, que describen las estructuras y contenidos de los archivos de almacenamiento y control del sistema.*
 - *./ejecutable -v [-e árbol de Editorial, -a árbol de Autor, -t hash de Título, -p hash de Palabra] "Nombre Archivo"*
 - *Nombre y estructuras para los archivos generados:*
 - *Archivo de Estructura de control: "Nombre Archivo"_Índice, "Nombre Archivo"_tabla.*
 - *Archivos de control de espacios libres: "Nombre Archivo"_libres.*
 - *Archivos de bloques de datos: "Nombre Archivo"_datos.*
 - *Estructura: para árboles la indicada en teórica. Separadores: Bloques con "|", Registros con ";", atributos con ",".*

3.2 Documentación

- *General*
 - *Diagrama de clases o módulos (según corresponda)*
 - *Especificación de cada clase o módulo (según corresponda)*
 - *Diagramas de secuencia o intercambio de mensajes entre capas. Mostrar escenarios.*
 - *Planificación (identificación de tareas, estimación de duración y asignación)*
 - *Bugs conocidos*

- Manual de usuario. Indicaciones generales del trabajo práctico, modo de instalación y ejemplos de uso.

- *Física – Organización*

- Organización de registros*

- ¿Cómo delimitan la longitud de un registro y de un campo variable?. Mostrar los campos que posee y cuanto espacio ocupa cada uno.
 - Indicar que información administrativa se utiliza.

- *Índices – Búsqueda*

- Hashing*

- Función de hashing utilizada. Criterio de elección.
 - Tamaño de Buckets.
 - ¿Cuál es el factor de empaquetamiento que utilizan?

- Árbol B+*

- *Sequence set*: ¿Cómo delimitan la longitud de un registro y de un campo variable?. Mostrar los campos que posee y cuanto espacio ocupa cada uno. Política utilizada para split / concatenación de bloques. ¿Hacen algún tipo de redistribución?
 - *Index set*: Mostrar los campos que posee y cuanto espacio ocupa cada uno. ¿Cómo se generan los separadores? ¿Cómo se eliminan separadores? Indicar que condiciones deben surgir en cada caso. ¿Existe algún tipo de concatenación de separadores dentro de la página? ¿Qué método se utiliza para la búsqueda de un separador dentro de la página? (binaria, secuencial, otro).

4 Referencias

Folk, Michael. Zoellick, Hill. *File Structures*. 724 páginas.

Smith, Peter. Barnes, G. *Files and Databases: An Introduction*. Addison-Wesley.