

CD UJI LIST-SERV (SERVIDOR DE CORREO ELECTRÓNICO DE LA BIBLIOTECA DE LA UNIVERSIDAD JAUME I PARA LA DISTRIBUCIÓN DE LOS SUMARIOS DE LAS PUBLICACIONES PERIÓDICAS)

Fernando Riera Candell

Estudiante de 4º curso de Ingeniería en Informática en la Universidad Jaume I

Resumen

La aplicación CD UJI LIST-SERV permite a los usuarios de la Biblioteca de la Universidad la consulta remota de los sumarios pertenecientes a las publicaciones especializadas que recibe periódicamente esta institución.

Para el desarrollo y puesta en funcionamiento de la aplicación se ha seguido la filosofía de procesamiento Cliente-Servidor. La aplicación actúa a modo de Servidor, respondiendo a todas las peticiones de servicio solicitadas por los distintos usuarios que serán los clientes de dicha aplicación.

Por otra parte, esta aplicación soporta toda una serie de operaciones sobre los sumarios de las publicaciones que se realizan a través del intercambio de Correo Electrónico (e-mail) con el Servidor CD UJI LIST-SERV.

Funcionamiento del servicio de distribución de sumarios de la biblioteca de la UJI

Para comprender mejor la necesidad que nos ha llevado a la creación de esta aplicación estudiaremos en primer lugar cual es el funcionamiento habitual del servicio de distribución de sumarios de la biblioteca de la UJI.

La biblioteca de la Universidad Jaume I recibe periódicamente los ejemplares de las publicaciones periódicas a las que se encuentra suscrita.

El método que se seguía era hacer fotocopias de los sumarios de estas publicaciones y repartirlas por los distintos departamentos de la Universidad.

De esta forma los investigadores de cada departamento conocían el contenido de cada una de estas publicaciones sin tener que desplazarse a la biblioteca, salvo que les interesase alguno de los temas tratados en las mismas y desearan consultar el ejemplar original.

Este método habitual en muchas bibliotecas tiene algunos inconvenientes que han motivado el que se haya creado una aplicación informática que realice automáticamente esta labor.

Entre estos inconvenientes figuran los siguientes :

- Es un método económicamente costoso, ya que el número de fotocopias que se hace es elevado.
- Es costoso en cuanto a horas de trabajo requeridas, ya que se necesita la dedicación de una persona para realizar esta labor.

- Se puede dar el caso de que no todos los investigadores tengan acceso a las fotocopias de los sumarios de las publicaciones por diversas razones.
- Se produce algún desfase de tiempo entre la recepción de la publicación y la obtención del sumario correspondiente por parte del investigador.

Por todo ello se pensó en la posibilidad de crear una aplicación informática que se encargara de realizar automáticamente esta labor.

Diseño de una posible solución

Después de un estudio detallado de las necesidades a las que debía dar respuesta un servicio automatizado de distribución de sumarios, se diseñó un Sistema que contempla las siguientes fases:

- 1) Obtención del sumario de una publicación por medio de un **scanner**, de modo que dispondríamos de un fichero en cuyo interior estaría el sumario de la publicación con el formato de salida del scanner, generalmente un **bit-map** (mapa de bits).
- 2) Pasar este fichero por un **O.C.R** (Optical Character Recognition) de modo que obtendríamos un fichero ASCII con el sumario.
- 3) Introducir este fichero de texto en la estación de trabajo, a través de **FTP** (File Transfer Protocol).
- 4) Distribuir los sumarios de las publicaciones a todos los usuarios de la Biblioteca de la Universidad a través del correo electrónico. Para este punto sería necesario implementar una aplicación que se encargara de realizar esta labor. (La aplicación ha sido implementada en C y se le ha llamado CD UJI LIST-SERV)

Estudio de la tecnología O.C.R. (Optical Character Recognition)

Todos los paquetes de software O.C.R. son capaces de leer una página escaneada y comprender "*virtualmente*" todos los caracteres que hay en esa página, permitiendo al usuario introducir el texto de dicha página en su procesador de textos, hoja de cálculo o cualquier otra aplicación.

Todos los sistemas O.C.R. trabajan con una imagen en formato bit-map (mapa de bits) obtenida a través de un modem-fax o un scanner.

El primer paso en el proceso de reconocimiento óptico de caracteres consiste en dividir la página en bloques de texto, y posteriormente esos bloques de texto son descompuestos en marcas individuales que corresponden a las letras. Estas marcas correspondientes a las letras son agrupadas en palabras, las palabras en líneas y las líneas con las mismas características, como justificación, tamaño, tipo de letra, son agrupadas en forma de párrafos.

El paso final del algoritmo consiste en determinar el orden de los párrafos en la página. Esto es muy fácil si tenemos un texto escrito en una sola columna, sin embargo, el proceso se hace mas difícil cuando se trata de un texto con varias columnas como por ejemplo la página de un periódico.

Las tablas, así como la distinción entre texto y dibujos pueden ser problemas particulares que requieren sus propios algoritmos de reconocimiento.

Para la conversión de las marcas individuales en caracteres se utilizan principalmente dos métodos:

— **Métodos basados en plantillas.** Los métodos basados en plantillas mantienen una colección de modelos de todas las posibles letras en todas las posibles fuentes. Las marcas que corresponden a las letras son comparadas con estas plantillas. Aquella plantilla que más se asemeje a la marca será la letra en cuestión a la cual corresponde la marca.

— **Métodos basados en los rasgos de las letras.** Estos métodos tratan de descomponer una letra en un conjunto de rasgos que corresponden a las características de dicha letra. Por ejemplo, la letra T puede ser considerada para este tipo de métodos como una raya vertical cruzada en su parte superior por una raya horizontal.

Tasa de acierto de los O.C.R.

Para medir la tasa de acierto se ha escaneado el mismo documento obtenido a través de distintos dispositivos. Luego se ha pasado cada uno de los documentos a través de un O.C.R. y se ha contado el número de fallos en el proceso de reconocimiento para cada uno de estos documentos. Los resultados obtenidos son los siguientes:

Dispositivo	Tasa de acierto	Errores por página
Impresora Daisy-wheel	99,4%	9 errores
Fax	93,1%	103 errores
Fotocopiadora	91,1%	133 errores
Impresora de agujas	92,1%	118 errores
Impresora láser	99,9%	1 error

NOTA: Considerando que hay 1.500 caracteres en una página

CD UJI LIST-SERV. (Servidor de Correo Electrónico)

Gracias a esta aplicación implementada en C los usuarios de la Universidad Jaime I tienen acceso a los sumarios de las publicaciones que son recibidas periódicamente por la Biblioteca de la Universidad.

El usuario puede solicitar una serie de operaciones sobre los sumarios de las publicaciones enviando mensajes a través del correo electrónico. Estos mensajes son enviados a la dirección de correo donde se encuentra instalada esta aplicación.

Las operaciones se solicitan incluyendo en el cuerpo del mensaje que se envía los comandos reservados que son reconocidos por una **función** que los interpreta y realiza las acciones pertinentes contestando a los Mails recibidos. En resumen, podemos ejecutar los siguientes comandos:

Comandos generales

Los comandos generales son aquellos comandos cuya ejecución puede ser solicitada por cualquier persona. Por lo tanto, siempre que llegue un mensaje que contenga alguno de estos comandos, será contestado sin realizar ninguna comprobación sobre la identidad de la persona que ha solicitado dicho servicio. Los comandos generales soportados por la aplicación son los siguientes:

- **help.** El usuario que envía el mensaje recibe vía correo electrónico un fichero cuyo contenido es una explicación de cada uno de los comandos que puede ejecutar y como hacerlo.
- **suscribe+nombre_publicación.** El usuario que envía el mensaje será suscrito a la lista de usuarios de la publicación "nombre_publicación", de modo que cuando se mande el sumario de la publicación a los usuarios suscritos a la misma, se le enviará a él también a la dirección con la que está registrado. (La dirección de procedencia de un mensaje se encuentra en la cabecera del mismo, junto con la fecha y Subject del mensaje.)
- **remove+nombre_publicación.** El usuario será dado de baja de modo que ya no figurará en la lista de usuarios suscritos a esta publicación y por lo tanto ya no se le enviará periódicamente el sumario, sino que lo tendrá que solicitar expresamente cada vez que desee obtener uno de ellos.
- **get_inde+nombre_publicación.** Con este comando el usuario solicita el sumario actual de una publicación dada sin la necesidad de suscribirse a la lista de usuarios de dicha publicación.
- **search -a + palabra_clave.** (-a significa en los actuales). Se busca en los ficheros correspondientes a los sumarios actuales de las publicaciones esa palabra clave, de modo que se envía al usuario los sumarios de las publicaciones que contengan esta palabra clave.
- **search -s+ palabra_clave+publicación.** (-s significa solo). Se busca la palabra clave en todos los sumarios almacenados de una determinada publicación.
- **getdir -a.** (a significa actuales). La aplicación configura un fichero cuyo contenido son los nombres de las publicaciones cuyo sumario actual se encuentra almacenado.

Este fichero es enviado al usuario que lo solicita, de modo que obtiene un directorio de las publicaciones actuales.

- **getdir -s +nombre_publicación**. El usuario recibe igualmente un fichero cuyo contenido es una lista de los sumarios que se encuentran almacenados de una publicación dada.

Comandos restringidos

A los Administradores de la aplicación se les permite ejecutar una serie de operaciones que no son accesibles al resto de los usuarios. Se trata de los comandos restringidos. Por lo tanto, cada vez que se recibe un mensaje que solicita la ejecución de un comando restringido, se comprueba que la persona que solicita el servicio tiene permiso para la ejecución de dicho comando. (Todos los nombres de personas con privilegios especiales se encuentran almacenados en un fichero llamado **privilegiados**).

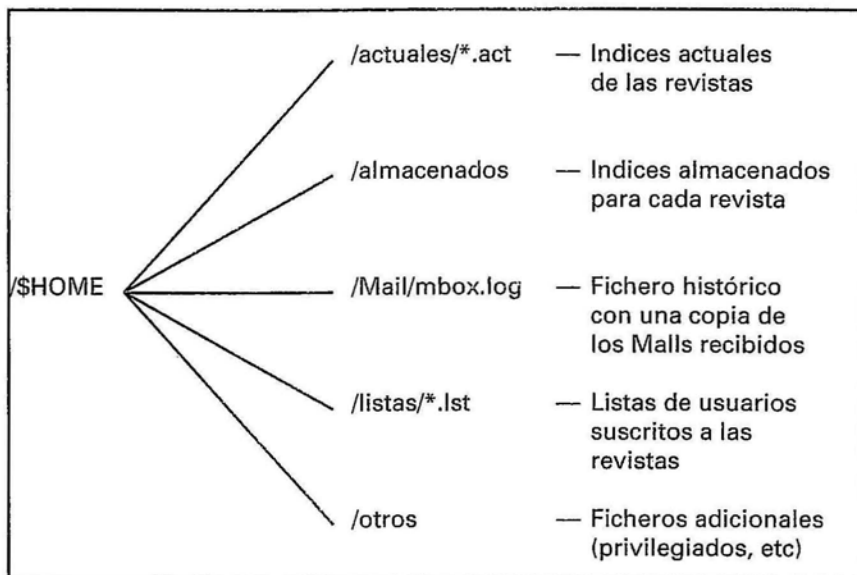
Por otra parte, existe un fichero llamado **mbox.log** que se encuentra en la ruta **/\$HOME/Mail/mbox.log** en el que la aplicación almacenan todos los mensajes que han sido tratados y ejecutados. Este fichero puede ser consultado por los administradores de modo que conocerán cualquier operación solicitada por los usuarios.

Los comandos restringidos permiten realizar las siguientes operaciones:

- **actualiz+nombre_publicación**. El resto del cuerpo del mensaje será el sumario de la publicación. Con esta operación actualizaremos el sumario actual de una publicación de modo que el que hasta ese momento era actual pasará a ser almacenado con el nombre que le corresponda. (Cuando el sumario de una publicación deja de ser el actual y se almacena, se guarda con el nombre de la publicación y con la extensión **.dia.mes.año**, siendo día, mes y año la fecha en la que ese sumario se almacena). Sin embargo, esta no es la única forma de actualizar el sumario de una publicación, sino que también puede hacerse introduciendo el fichero con el sumario a través del FTP.
- **rm_mbox**. El fichero temporal que contiene la memoria de todas las operaciones que han sido ejecutadas por la aplicación es inicializado. (Para prevenir el que ese fichero adquiera un gran tamaño).
- **ver_mbox**. El administrador recibe en su cuenta propia un mensaje cuyo contenido es el fichero histórico **mbox.log**. De esta forma puede consultar las operaciones que se han realizado.

Estructura de almacenamiento

La figura siguiente representa la estructura de almacenamiento para los sumarios de las publicaciones, listas de usuarios y todos los programas y ficheros relacionados con el funcionamiento de la aplicación CD UJI LIST-SERV.



El directorio `/listas/nombre_publicacion.lst` contiene los ficheros con los nombres de los usuarios suscritos a cada publicación. Para cada publicación que gestione esta aplicación (en la biblioteca son mas de 1000 los títulos que se reciben) tendremos un fichero que contendrá los nombres de los usuarios suscritos a la misma. Cada vez que se reciba una nueva publicación, su sumario será enviado a todos y cada uno de los nombres (direcciones de correo electrónico) que figuran en el fichero de usuarios suscritos a esa publicación. El nombre de cada uno de los ficheros corresponde al nombre de la publicación seguido de la extensión .lst (lista).

En el directorio `/almacenados/nombre_public/nombre_public.fecha` se encuentran los sumarios almacenados de cada publicación. Para cada publicación tenemos un subdirectorio dentro del directorio almacenados que se llama como la publicación de la cual contiene los sumarios. Cuando se almacena un sumario, se guarda con el nombre de la publicación seguido de la extensión fecha, siendo fecha el día, mes y año en el que ese sumario es almacenado.

(Toda la gestión y almacenamiento de los sumarios es realizada por la aplicación CD UJI LIST-SERV de forma automática cada vez que se introduce un nuevo sumario.)

El fichero `/Mail/mbox.log` contiene una memoria de todas las operaciones solicitadas por los clientes de la aplicación. De esta forma, los administradores pueden consultar este fichero para ver que operaciones han sido solicitadas y atendidas.

Facilidades para el usuario

El paquete CD UJI LIST-SERV incluye también un programa que facilita la labor al usuario, de modo que el proceso de confección del mensaje es transparente al mismo.

El menú general que aparece al ejecutar el programa es el siguiente:

1. Envío de mensaje de suscripción a una publicación.
 2. Envío de mensaje de borrado de una publicación.
 3. Envío de mensaje de solicitud del sumario actual de una publicación.
 4. Envío de mensaje de búsqueda de palabra clave.
 5. Envío de mensaje de obtención de directorios de publicaciones disponibles.
 6. Envío de mensaje de solicitud de ayuda.
 7. Modificación de la dirección de la aplicación list_serv.
4. Envío de mensaje de búsqueda de palabra clave.
 - 4.1. Buscar en los sumarios actuales.
 - 4.2. Buscar en los sumarios almacenados.
 5. Envío de mensaje de obtención de directorio de los sumarios de las publicaciones disponibles.
 - 5.1. Obtener el directorio de las publicaciones actuales.
 - 5.2. Obtener directorio de los sumarios almacenados de una publicación.

Si el usuario instala esta aplicación en su cuenta puede confeccionar y enviar los mensajes sin necesidad de conocer cual es la sintaxis de cada uno de ellos, ya que el programa se encarga de hacerlo todo por él. Simplemente el programa pide al usuario los datos que necesite y una vez tiene configurado el mensaje se lo hace saber al usuario y espera el permiso del mismo para enviar el Mail.

Facilidades para el administrador

Al igual que el usuario dispone de un programa que le permite generar automáticamente los mensajes, el administrador dispone de una aplicación que le permite realizar directamente todas las acciones que podría realizar a través de Mails.

De esta forma el administrador tiene un control constante e inmediato sin la necesidad de obtener esta información mediante el intercambio de correo electrónico con la aplicación CD UJI LIST-SERV.

El menú general que aparece al cargar esta aplicación es el siguiente :

1. Suscribir a una persona en una lista.
2. Ver el contenido de una determinada lista.
3. Borrar a una persona de una lista.
4. Enviar el sumario de una publicación.
5. Opciones sobre el fichero histórico.
6. Ver publicaciones disponibles o archivadas.
7. Personas con privilegios especiales.
8. Salir.
5. Opciones sobre el fichero histórico.
 - 5.1. Ver el contenido del fichero histórico.
 - 5.2. Inicializar el fichero histórico.

6. Ver publicaciones disponibles o archivadas.
 - 6.1. Obtener un directorio de las publicaciones actuales.
 - 6.2. Obtener un directorio de los sumarios almacenados de una publicación.
7. Personas con privilegios especiales.
 - 7.1. Ver listado de personas con privilegios.
 - 7.2. Expandir privilegios a una persona.
 - 7.3. Revocar privilegios a una persona.

Puesta en ejecución de la aplicación

Un punto de especial relevancia en el momento de implementar la aplicación es estudiar la forma en la que se va a poner en ejecución cada uno de los programas que forman parte del paquete CD UJI LIST-SERV.

Las posibilidades son las siguientes:

- 1) Incorporar estos programas en el fichero **.profile** de la cuenta de la estación de trabajo UNIX donde está instalada la aplicación. De esta forma, cada vez que se entre en la cuenta se ejecutarán estos programas y los mensajes de solicitud de servicio recibidos desde la última conexión serán tratados por la aplicación.
- 2) Otra opción es utilizar el comando **"at"**, de modo que especificamos a que hora queremos que se ejecute la aplicación. Por ejemplo, podríamos ejecutar la aplicación a las 12:00 de la mañana y a las 12:00 de la noche todos los días.
- 3) Por último, tenemos la opción de ejecutar los programas que forman parte del paquete cada vez que llegue un mensaje a la cuenta donde está instalada la aplicación. Para esto, debemos modificar el fichero **.forward** que contiene las órdenes que se ejecutan cada vez que llega un mensaje a la cuenta. Esta parece ser la mejor solución de todas ellas, ya que obtenemos una respuesta inmediata por parte del Servidor.

En cualquier caso esta decisión depende de los usuarios de la aplicación. Serán ellos los que decidan la forma de ejecución del paquete de acuerdo con sus necesidades particulares.

Conclusiones

Gracias a esta aplicación, la Biblioteca de la Universidad Jaume I ha puesto a disposición de sus usuarios un método para que estos últimos puedan conocer el contenido de cada una de las publicaciones que se reciben.

Las ventajas que aporta esta aplicación informática son:

- 1) No es necesario personarse en la Biblioteca para obtener información de los últimos artículos que aparecen en las publicaciones recibidas por la Universidad, sino que el investigador desde su despacho o el alumno desde las aulas de informática tendrán acceso inmediato a toda la información de este tipo que requieran.
- 2) La aplicación supone un ahorro significativo de dinero y se trata de un método lim-

pio y completamente ecológico, dado que no se necesita fotocopiar los sumarios de las publicaciones. Por otra parte, supone un ahorro de horas de trabajo por parte del personal de la Biblioteca de la Universidad.

- 3) Gracias a la estructura que se ha seguido en la implementación de la aplicación, esta puede ser ampliada en el futuro con el objetivo de satisfacer las nuevas necesidades que surjan en la Biblioteca de la Universidad.
- 4) El control estadístico sobre la tasa de consultas que se realizan sobre cada publicación, realizado a mano en la actualidad, puede ser realizado automáticamente con una pequeña ampliación del paquete CD UJI LIST-SERV.

Bibliografía

SCHILD, Herbert : *Turbo C/C++*, Manual de Referencia. McGraw-Hill, 1989.

ROCHKIND, Marc : *Advanced Unix Programming*. Prentice-Hall, 1990.

WAYNER, Peter : «Optimal Character Recognition». Byte. Diciembre 1993. Pag 203-210.

BARKER, D. : «Typereader takes O.C.R. toward better Recognition». Byte. Diciembre 1992. Pag 213-215.