

SISTEMAS LEGADOS

Práctica 3

Objetivo:

- Implementar una página web que, con la ayuda de un *wrapper* que genere la entrada y analice la salida, acceda a una aplicación legada BASIC/MS-DOS con interfaz de texto.

Descripción de sistema legado

La aplicación legada es un programa en BASIC para MS-DOS, desarrollada en un IBM PC a comienzos de la década de 1980, que almacena en fichero la información de distintos programas para el ordenador Sinclair ZX Spectrum (nombre, tipo de juego, y cinta/s donde se encuentra almacenado). Se pueden realizar varias operaciones (ver Figura 1) entre las cuales se encuentra la ordenación según distintos campos, el listado de los programas según dicho orden, y la búsqueda por nombre de un programa concreto.

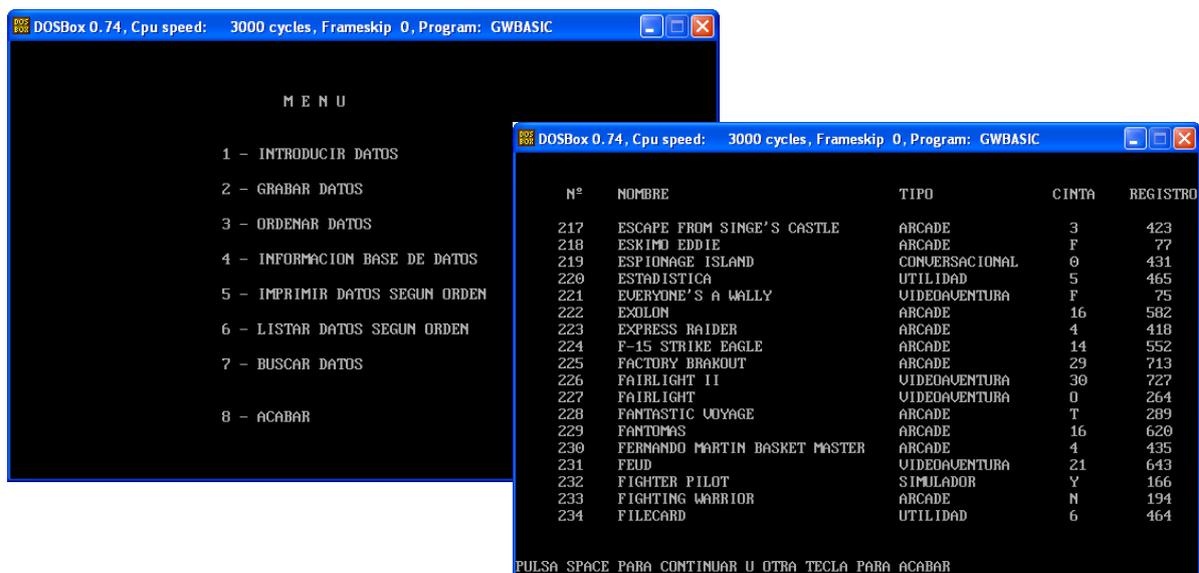


Figura 1: La aplicación MS-DOS/BASIC que tenemos que encapsular

Puedes descargar la aplicación de la página web de la asignatura. Descomprime `Database-MSDOS.rar` y `database.bat` lanzará la aplicación usando el emulador de MS-DOS DOSBox (incluido); la aplicación legada está en la carpeta `Database` y no debe cambiarse nada de lo que contiene, hay que encapsular su uso.

Objetivo

Crear una página web que permita acceder a los datos manejados por dicha aplicación desde un ordenador remoto a través de un navegador web. Concretamente, la página web debe permitir las siguientes funcionalidades:

- Visualizar en todo momento el número de registros del sistema legado.
- Dado el nombre de un programa, listar sus datos.
- Dado el identificador de una cinta (A-Z, 0-n), listar los programas que contiene en el orden en que se encuentran grabados. NOTA: un programa puede estar en más de una cinta (ej., para un programa que esté en la cinta A y en la 45 → cinta="A-45").

Dicha página web recurrirá a los servicios de un *wrapper* que encapsulará la aplicación legada, generando la entrada correspondiente a la operación solicitada por el usuario desde el navegador, y capturando y analizando la salida en pantalla para obtener los datos que se mostrarán en el navegador como respuesta. Posiblemente alguna operación individual a través de la página web requerirá, de forma transparente para el usuario, de varias interacciones con la aplicación legada.

SISTEMAS LEGADOS

Generación de la entrada de la aplicación legada

Para que el wrapper pueda generar la entrada para la aplicación legada podemos usar algo parecido a la clase `Robot` de Java (`java.awt.Robot`) que permite generar eventos de bajo nivel de teclado y ratón. Un ejemplo es el siguiente:

```
Robot robot = new Robot();
robot.keyPress(KeyEvent.VK_H); robot.keyPress(KeyEvent.VK_O);
robot.keyPress(KeyEvent.VK_L); robot.keyPress(KeyEvent.VK_A);
```

Captura de la salida de la aplicación legada

En este contexto, una solución para que el wrapper pueda acceder a la salida de la aplicación legada es capturar gráficamente la imagen de la pantalla y hacer un reconocimiento óptico de caracteres (OCR) de la misma. Para la primera tarea puede utilizarse de nuevo la clase Java `Robot`, como se ve en el ejemplo siguiente:

```
// Obtención de las coordenadas de pantalla x1, x2, y1, y2 de la aplicación legada
...
BufferedImage capture =
    new Robot().createScreenCapture(new Rectangle(x1, y1, x2 - x1, y2 - y1));
```

Para el OCR podemos utilizar la aplicación de software libre `Tesseract-ocr` [6], para la cual existe un wrapper en Java denominado `Tess4J` [7]. Esta aplicación puede entrenarse para reconocer distintos idiomas en distintas fuentes. En nuestro caso, se ha dejado en la web de la asignatura un fichero (`spa.traineddata`) con datos de entrenamiento ya generados para reconocer el tipo de texto utilizado por la aplicación legada. Este fichero debe ser copiado en la carpeta `tessdata` de la aplicación `Tess4J`. El ejemplo siguiente muestra como invocar el OCR que utilizaremos:

```
Tesseract1 ocr = new Tesseract1();
ocr.setLanguage("spa");
String result = ocr.doOCR(capture);

// Análisis de la salida
...
```

Es conocido que el comportamiento de este programa OCR para esta tarea no es el correcto al 100%. Para mejorar dicho comportamiento podemos bien entrenar mejor al OCR (ver documentación [6]), o bien utilizar cualquier otro OCR disponible en la Web.

NOTA: Si a pesar de un mejor entrenamiento los resultados por OCR no son satisfactorios al 100%, puede utilizarse *cualquier* otro método o software (todo vale excepto cambiar la aplicación legada en BASIC o dejar de usarla íntegramente) para capturar la salida de la aplicación legada y que el resultado de su encapsulación sea lo más parecido posible a la ejecución directa de dicha aplicación.

Documentación

Redactar 4-5 páginas aprox. detallando los pasos realizados para hacer la encapsulación de la aplicación legada, incluyendo aquellos que no han servido para llegar a la solución final, las principales dificultades encontradas, el resultado al utilizar las distintas técnicas o configuraciones del OCR, pantallazos/video que muestre el funcionamiento de la aplicación encapsulada, el reparto de tareas entre los miembros del grupo, el tiempo real invertido en cada tarea, y cualquier otra información que se considere de interés, incluyendo todo lo necesario para poder ejecutar las practicas desde Windows haciendo doble click en un fichero.

Referencias

[6] <https://code.google.com/p/tesseract-ocr/>

[7] <http://tess4j.sourceforge.net/>