



**Universidad de Valladolid**

## **Escuela de Ingeniería Informática**

**TRABAJO FIN DE GRADO**

Grado en Ingeniería Informática  
(Mención Computación)

---

# **Procesado de lenguaje natural mediante Redes Neuronales Recurrentes.**

**Manual de usuario**

---

Autor:

**Esther Cuervo Fernández**

Tutores:

**Teodoro Calonge Cano**

**Joaquín Adiego Rodríguez**

## 1. Inicio

Tras ejecutar la aplicación, se mostrará la pantalla inicial, donde podrá ver los documentos en el sistema, si existe alguno, y un menú desde el que se pueden lanzar los distintos casos de uso.

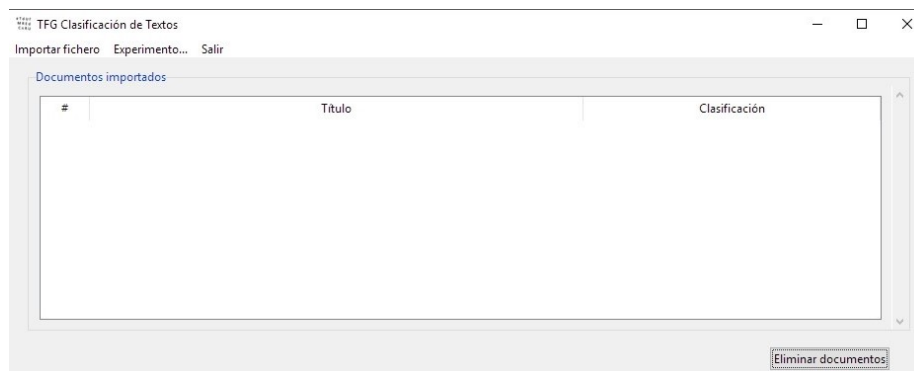


Figura 1: Pantalla inicial de la aplicación

Si se seleccionan uno o más documentos de la lista y se pulsa el botón de Eliminar documentos, se mostrará un mensaje de confirmación:

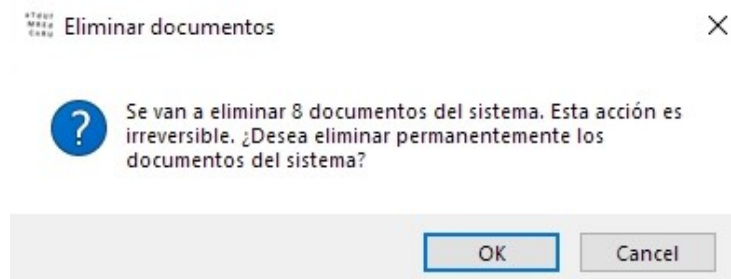


Figura 2: Pantalla de confirmación de la eliminación.

Si se confirma, se eliminarán los documentos del sistema y se actualizará la pantalla principal.

También es posible visualizar el contenido de un documento, haciendo doble click sobre uno de los documentos en la lista, lo que hará surgir una pantalla como esta:

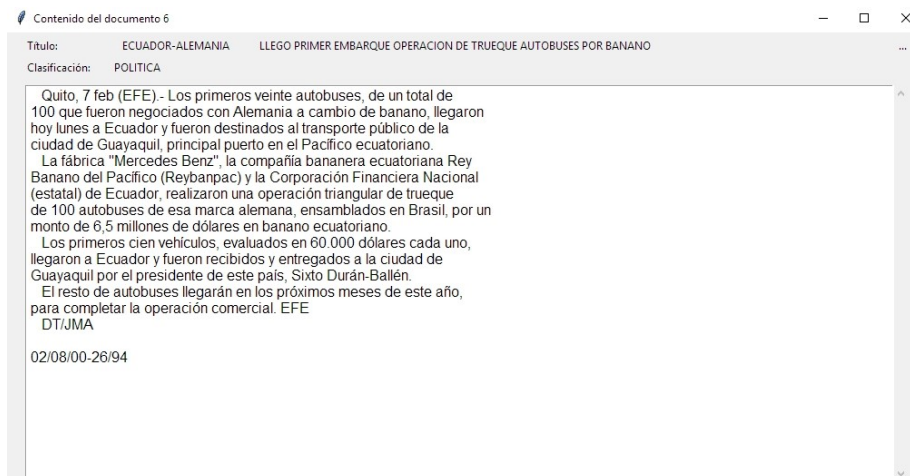


Figura 3: Pantalla de visualización de un documento desde la pantalla principal.

## 2. Importar documentos

Desde esta pantalla podrá acceder a la importación de ficheros, pulsando sobre la primera opción del menú, **Importar fichero**. Esto abrirá una pantalla donde puede seleccionar los ficheros a importar.

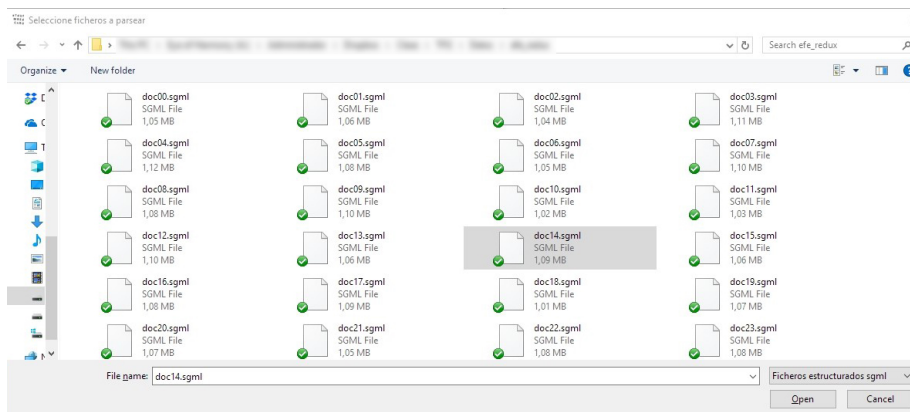


Figura 4: Primera pantalla de la importación de ficheros.

Los documentos estructurados .sgml a importar deben tener la siguiente estructura:

```
<DOC>
...
<TITLE> Titulo </TITLE>
```

```

...
<CATEGORY>Clase</CATEGORY>
...
<TEXT>Cuerpo</TEXT>
...
</DOC>

```

Las etiquetas *TITLE* y *CATEGORY* son opcionales, y los ficheros pueden tener esta estructura varias veces, se detectará un documento por cada estructura. Si alguno de los ficheros seleccionados no sigue esta estructura, se mostrará el siguiente mensaje de error:

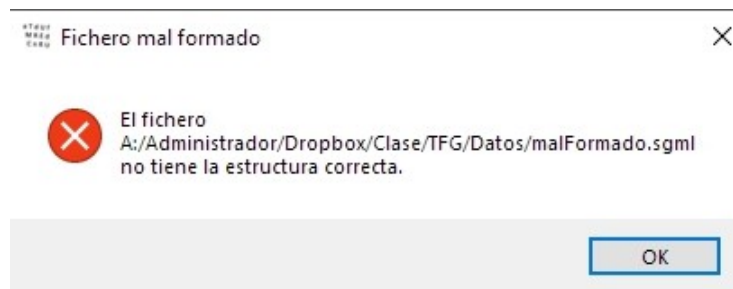


Figura 5: Pantalla de error en importado: Error en la estructura del fichero.

Puede seleccionar ficheros de texto plano seleccionando en la parte inferior el tipo de fichero a importar:

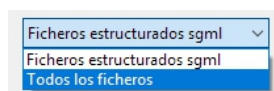


Figura 6: Pantalla de cambio del tipo de fichero a texto plano.

Si alguno de los ficheros no puede ser leído como fichero de texto (como un .pdf, .mp3, etc), se mostrará el siguiente mensaje de error:

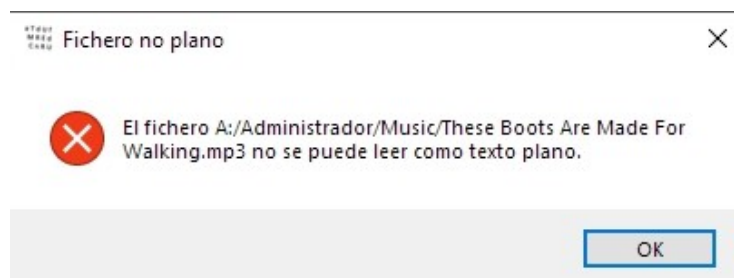


Figura 7: Pantalla de error en importado: El fichero no puede ser leído como fichero de texto.

Si todos los ficheros introducidos son correctos, se mostrará la pantalla para editar los atributos de los documentos parseados. Se puede seleccionar un documento y cambiar su título y clasificación, y el botón Guardar cambios cambia el documento.

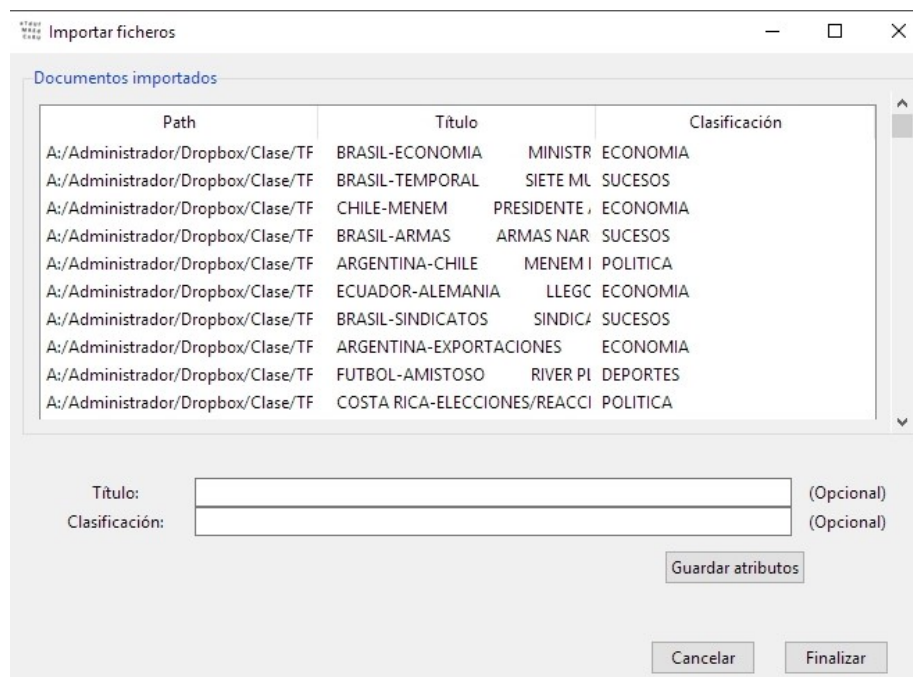


Figura 8: Pantalla de editado de atributos de documentos.

Desde esta pantalla se puede hacer doble click sobre uno de los documentos parseados, y se mostrará su contenido:

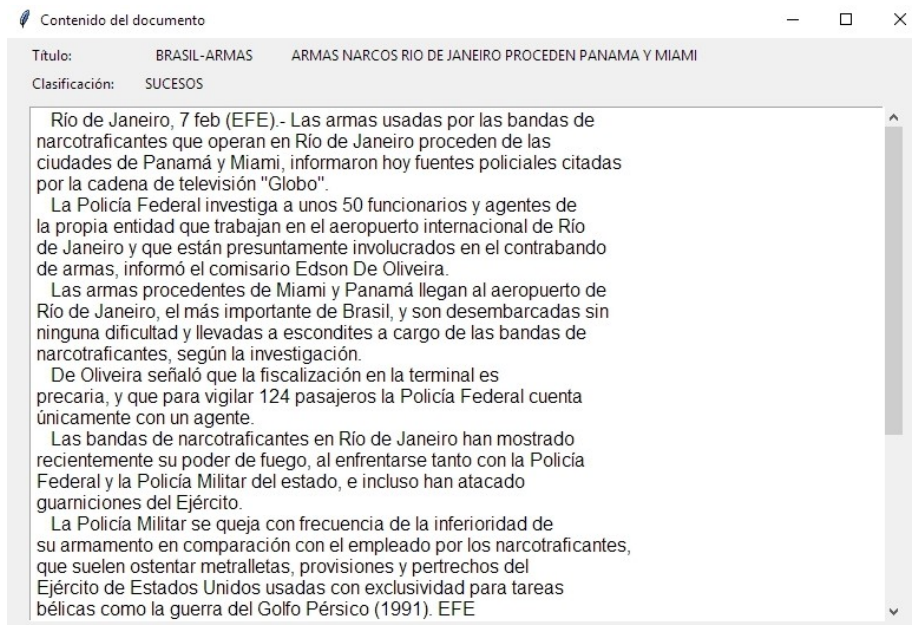


Figura 9: Pantalla de visualización de documentos desde el importado.

Al pulsar en Finalizar, se almacenan los documentos en la base de datos del sistema, y se vuelve a la pantalla principal, actualizada con los nuevos documentos.



Figura 10: Pantalla inicial de la aplicación actualizada

### 3. Entrenamiento

Para iniciar un experimento de entrenamiento, pulse sobre la opción Experimento... en el menú de la pantalla principal y seleccione Entrenamiento. Si no existen documentos con clasificación en el sistema se mostrará el siguiente

mensaje de error:

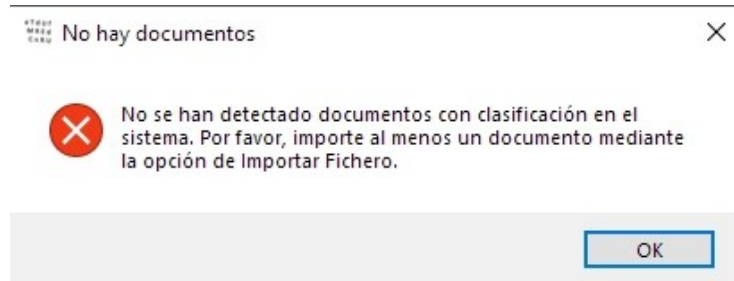


Figura 11: Pantalla de error en entrenamiento: No existen documentos con clasificación en el sistema.

Si existen documentos con clasificación, se mostrará la pantalla de selección de documentos y parámetros:

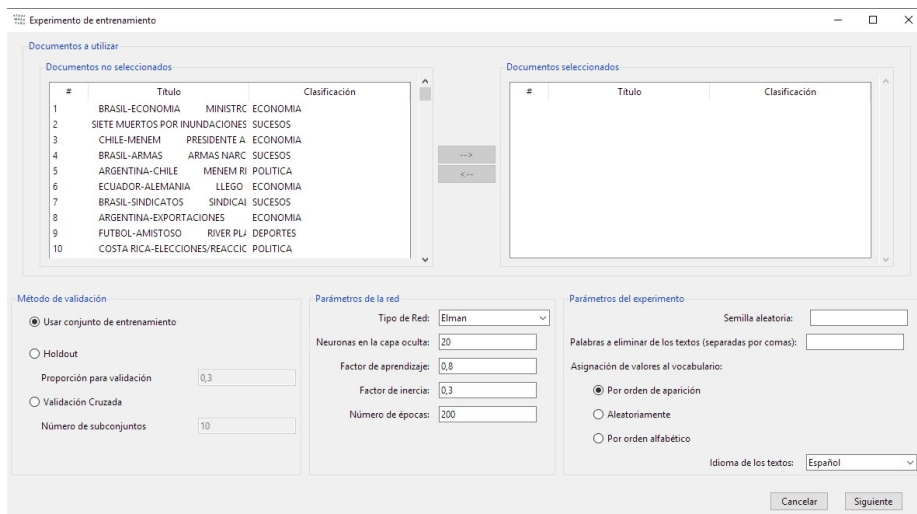


Figura 12: Pantalla de selección de documentos y parámetros.

La definición de cada parámetro es:

- **Método de validación.** Define la metodología para seleccionar el conjunto de validación entre los documentos seleccionados.
- **Tipo de red.** Tipo de red neuronal recurrente a utilizar.
- **Neuronas en la capa oculta.** Número de neuronas a utilizar para la capa oculta de la red.

- **Factor de aprendizaje.** Factor por el que se multiplica la variación de peso causada por el error cometido al entrenar una instancia.
- **Factor de inercia.** Factor por el que se multiplica la variación de peso causada por la instancia anterior.
- **Número de épocas.** Número de pasadas por todo el conjunto de entrenamiento a realizar.
- **Semilla aleatoria.** Opcional. Permite designar una semilla para los generadores de números aleatorios que utiliza la aplicación. Sirve para conseguir experimentos reproducibles. Debe ser un entero.
- **Palabras e liminar de los textos.** Opcional. Palabras que se tratarán como "palabras vacías.<sup>a</sup> la hora de procesar el texto. Se deben introducir separadas por comas.
- **Asignación de valores al vocabulario.** Permite indicar el orden en el que asignar los valores al vocabulario.

Si se introduce algún parámetro inválido, o el número de documentos y el método de validación seleccionado no son compatibles, se mostrará un mensaje de error. Si todos los parámetros y documentos son válidos, se inicia el entrenamiento, y se muestra la pantalla de progreso del experimento:

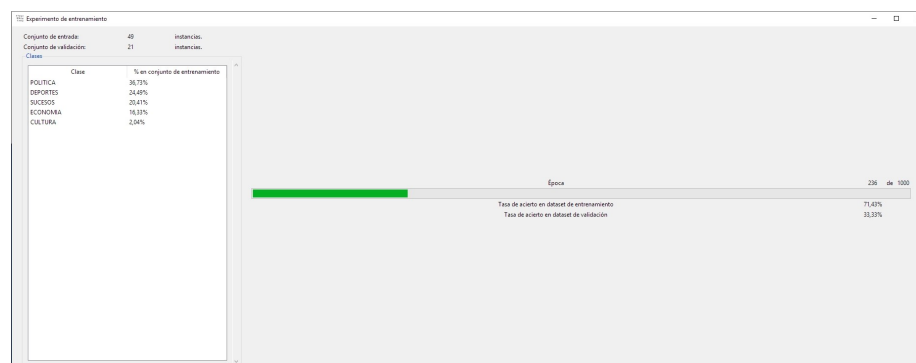


Figura 13: Pantalla de progreso del entrenamiento.

Esta pantalla muestra el tamaño de los conjuntos de entrenamiento y validación, y la distribución de clases en el conjunto de entrenamiento, así como la tasa de acierto en cada conjunto, que se actualiza cada época.

Al finalizar el experimento, se muestra la pantalla de resultados:



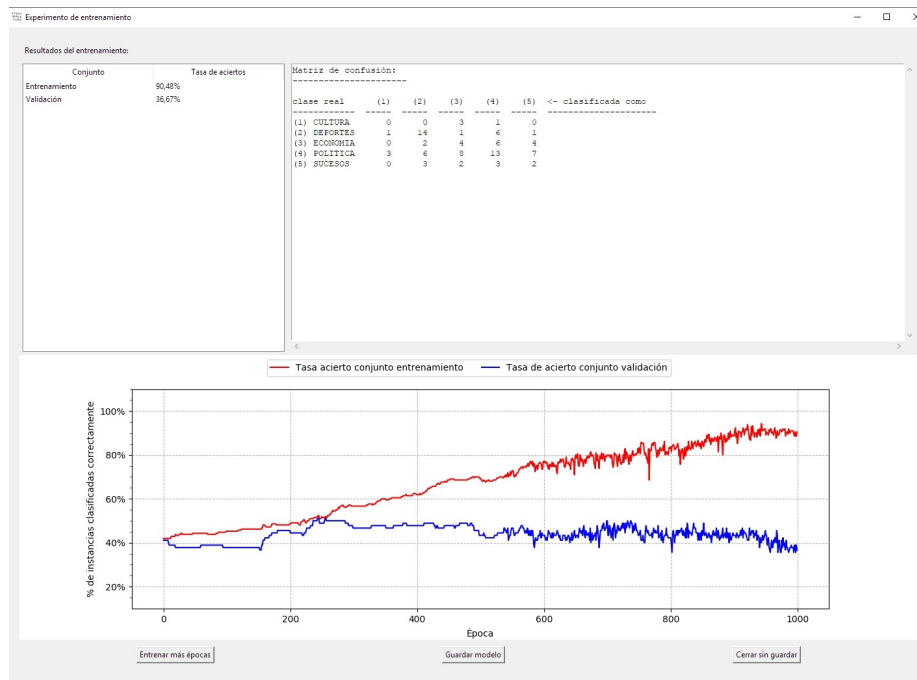


Figura 14: Pantalla de resultados del entrenamiento.

En esta pantalla se muestran las tasas de acierto finales en cada conjunto, la matriz de confusión, que muestra la relación entre clase real y la clase obtenida al finalizar el entrenamiento en el conjunto de validación. También se muestra un gráfico con la evolución a lo largo del entrenamiento de las tasas de acierto de cada conjunto.

El botón Entrenar más épocas permite reanudar el entrenamiento un número extra de épocas:

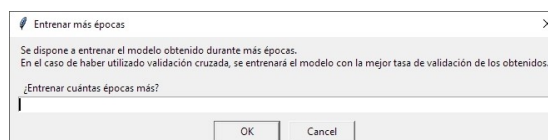


Figura 15: Pantalla de introducción de épocas extras.

Si el valor introducido no es un entero mayor que 0, se muestra un mensaje de error y se vuelve a la pantalla de resultado. Si el valor es válido, se vuelve a la pantalla de progreso.

El botón Guardar modelo permite guardar el modelo obtenido en el sistema con el nombre introducido:

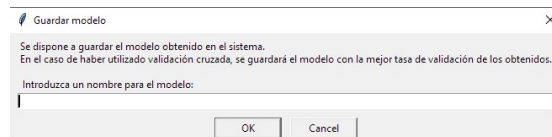


Figura 16: Pantalla de introducción de nombre del modelo a guardar.

Tras introducir un nombre la aplicación guardará el modelo y cerrará la pantalla de entrenamiento.

Si se selecciona la opción de Cerrar sin guardar se muestra un mensaje de confirmación y si se confirma se cierra la pantalla de entrenamiento.

## 4. Clasificación

Para iniciar un experimento de clasificación, pulse sobre la opción Experimento... en el menú de la pantalla principal y seleccione Clasificación. Si no existen documentos en el sistema se mostrará el siguiente mensaje de error:

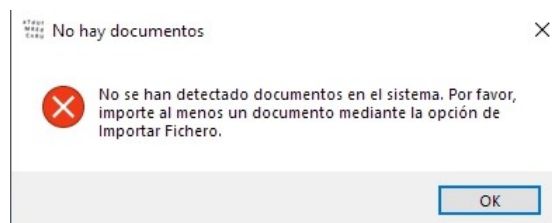


Figura 17: Pantalla de error en clasificación: No existen documentos.

Si existen documentos pero no modelos, se mostrará el siguiente mensaje de error:

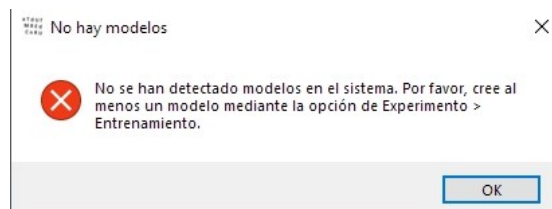


Figura 18: Pantalla de error en clasificación: No existen modelos.

Si existen documentos y modelos en el sistema, se mostrará la pantalla de selección de documentos y modelo:

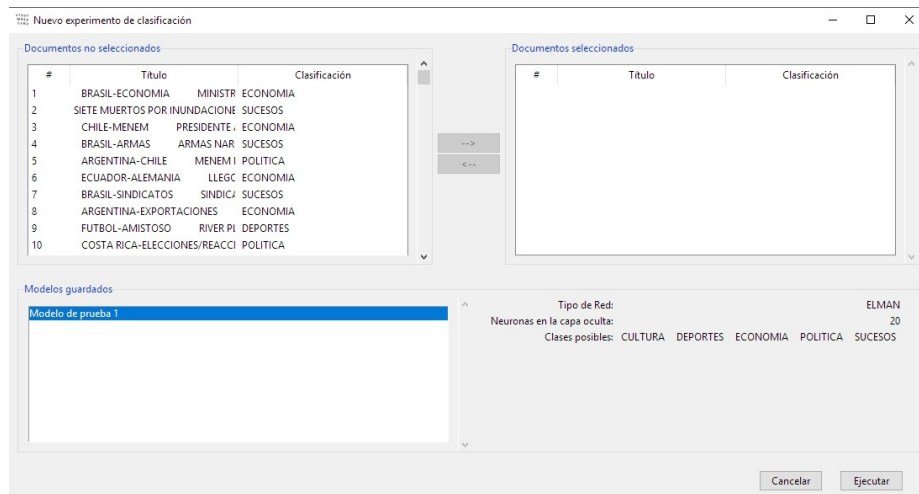


Figura 19: Pantalla de selección de documentos y modelo.

Al pulsar sobre cada modelo en la lista de modelos disponibles se actualizan los datos al modelo seleccionado.

Al pulsar Ejecutar se clasifican los documentos, y se muestra la pantalla de resultados:

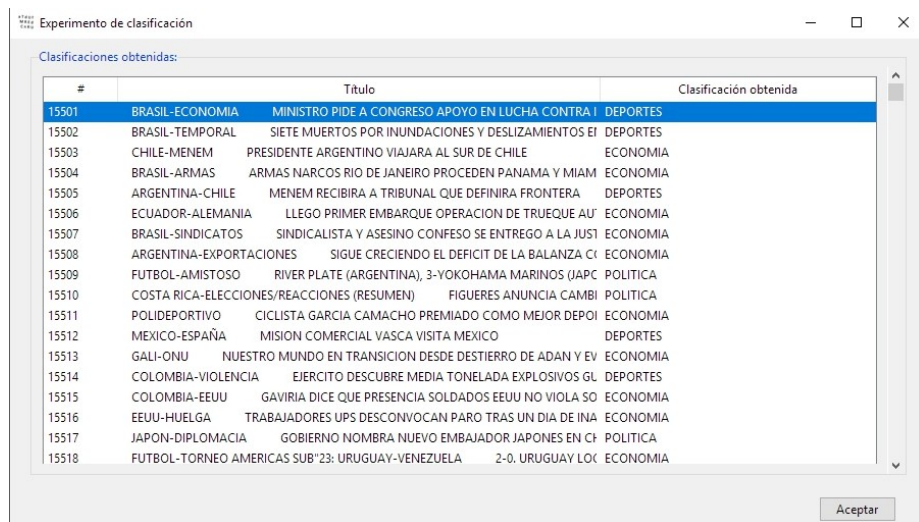


Figura 20: Pantalla de resultados de la clasificación.

En esta pantalla es posible hacer doble click en uno de los documentos, y se mostrará el contenido, título, y clasificación real de dicho documento:

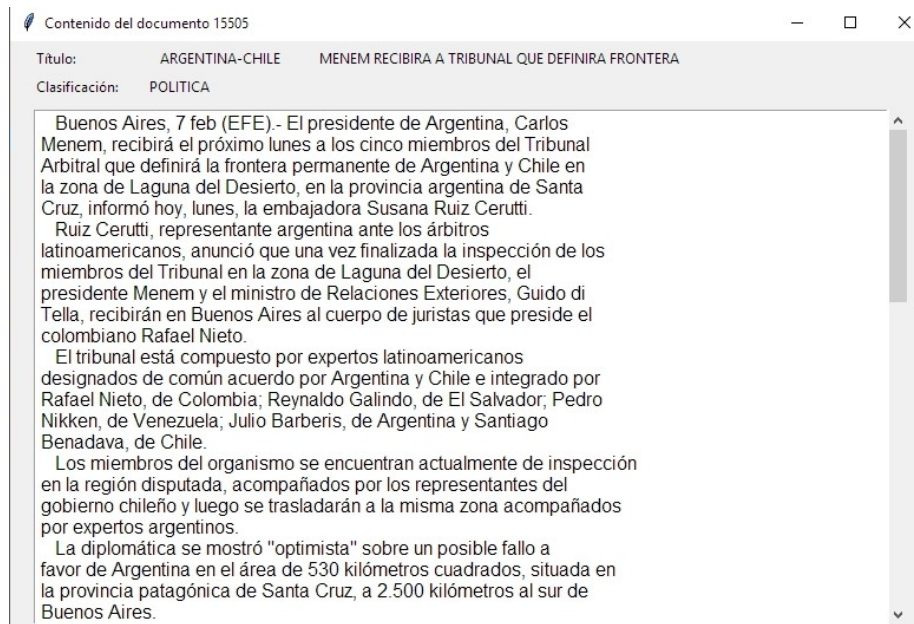


Figura 21: Pantalla de visualización del documento tras la clasificación.

Al pulsar en Aceptar en la pantalla de resultados se cierra la pantalla de clasificación.