

Título: Programa de lectura y almacenamiento de noticias del Marca

Descripción:

Este programa en Java permite obtener todas las noticias publicadas en los RSS de Marca, evitando duplicados, y almacenarlas en un fichero serializado para uso posterior. Utiliza DOM para parsear los archivos XML de cada feed RSS y controla que cada noticia se guarde solo una vez mediante su campo `guid`.

Funcionamiento paso a paso:

1. Definición de la clase Noticia:

- Se crea una clase `Noticia` con los campos: `title`, `link`, `description`, `guid`, `pubDate` y `categorias`.
- El campo `categorias` se implementa como un `ArrayList<String>`.
- La clase implementa `Serializable` para poder almacenar objetos en fichero.

2. Obtención de las URLs de RSS de Marca:

- Se lee la página `https://www.marca.com/rss.html` usando `URLConnection`.
- Se filtran solo las líneas que contienen enlaces a archivos `.xml`.
- Se extraen las URLs de los XML y se guardan en un `ArrayList<String>` para su posterior lectura.

3. Carga de noticias desde ficheros serializados previos:

- Al iniciar, el programa comprueba si existe el fichero `noticias_marca.ser`.
- Si existe, se carga el `ArrayList` de noticias previamente guardado, evitando volver a descargar noticias repetidas.

4. Lectura de cada RSS con DOM:

- Por cada URL de RSS, se parsea el XML usando DOM (`DocumentBuilder`).
- Se obtiene cada elemento `<item>` que representa una noticia.

- Se extraen los campos `title`, `link`, `description`, `guid`, `pubDate` y todas las `<category>`.

5. Control de duplicados:

- Antes de añadir cada noticia al ArrayList, se comprueba que su `guid` no exista ya en la lista.
- Esto evita que una misma noticia que aparezca en varios RSS se guarde múltiples veces.

6. Almacenamiento de noticias:

- Una vez recopiladas todas las noticias, se serializa el ArrayList completo en el fichero `noticias_marca.ser`.
- Esto permite cargar las noticias en ejecuciones futuras sin necesidad de volver a descargar los RSS.

Gestión de errores:

- Se controlan excepciones de conexión, parsing y serialización.
- Se muestra un mensaje si no se puede leer un XML concreto, pero el programa continúa con los demás RSS.

Aspectos técnicos:

- Se utiliza DOM para parsear los XML, cargando cada archivo en memoria.
- La serialización permite guardar y recuperar el estado completo de las noticias de manera eficiente.
- La estructura de control con `guid` garantiza la unicidad de cada noticia.
- El filtrado de URLs en la página de Marca asegura que solo se procesen archivos RSS válidos.

Resumen de flujo del programa:

1. Comprobar fichero serializado → cargar si existe.
2. Obtener URLs de RSS desde Marca.

3. Recorrer cada RSS y extraer noticias con DOM.
4. Evitar duplicados por `guid`.
5. Guardar todas las noticias en fichero serializado.
6. Mostrar mensajes de estado durante el proceso (noticias cargadas, errores, etc.).