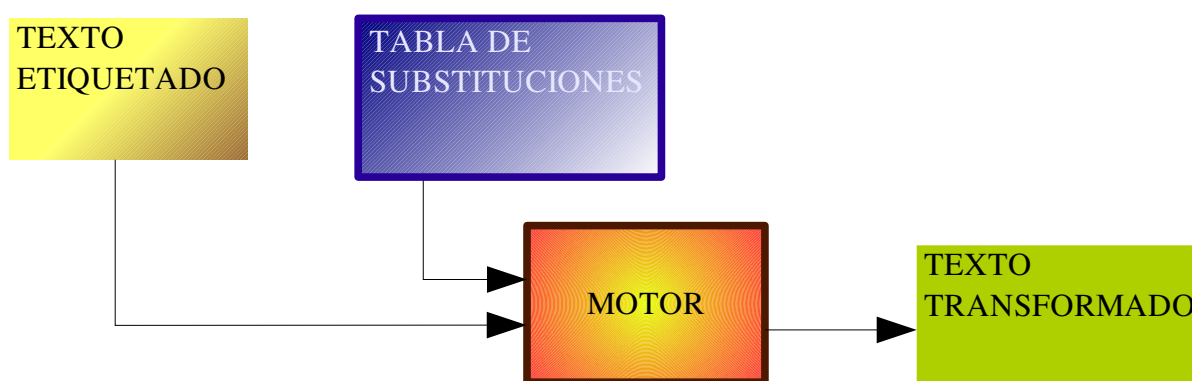


TDTxE: Motor Transformador de documentos (etiquetados)

Resumen

Este proyecto nace a partir de una necesidad concreta para convertir texto con etiquetas html4+css (con estilos) hacia html 3.2, debido a un problema de incompatibilidad con una aplicación que lo ha de manipular.

Esta necesidad se ve como una oportunidad para construir una aplicación más genérica que pueda transformar las etiquetas de cualquier documento de texto a partir de unas tablas de transformación que los propios usuarios puedan elaborar. El esquema de la aplicación sería el siguiente:



Los dos componentes principales son:

-**Motor de proceso:** que a partir de un texto de entrada le aplica una serie de substituciones.

-**Tabla de substituciones:** contiene la relación de todas las substituciones que se quieren aplicar con toda la información adicional que sea necesaria. Estas tablas han de poder ser fácilmente manipulables por el usuario para poder construir diferentes tablas de *conversión* y poder escoger en cada momento cual se quiere utilizar con un determinado texto.

También se necesitan 2 modos de funcionamiento:

-**Línea de ordenes:** recibe como parámetros los documentos de entrada, salida y tabla de substituciones. Pensado para poder hacer operaciones *batch* o llamadas desde otros procesos.

-**Interfaz gráfica:** Con una caja para escribir/enganchar texto y otra donde sale el texto procesado, una vez que se haya escogido también la tabla de substituciones a aplicar. Para utilizarla en casos en que se quiere hacer por ejemplo una "transformación rápida" a partir de fragmentos de un documento, texto residente en otra aplicación y en definitiva, casos en que no existe propiamente un fichero a convertir.

Utilizando las tablas adecuadas y con el mismo motor, se puede utilizar también para limpiar código, eliminar comentarios...

Motor de proceso

Como es habitual en los proyectos de software libre, partiremos de una **primera versión básica** que podrá evolucionar para incorporar opciones, mejoras de rendimiento, etc. también en función de lo que la propia comunidad demande y quiera aportar.

La versión inicial consistirá en un proceso iterativo: para cada entrada de la Tabla de Substituciones (TS a partir de ahora), se procesa todo el documento entrado. Cuando se llega al final de la TS y una vez procesada su última entrada, habrá finalizado el proceso y tendremos el fichero transformado. Por tanto, se ha de tener presente lo siguiente (para la primera versión):

- Las substituciones son textuales es decir, no se aplican patrones de cadenas.
- Solo funcionará en modo consola. El programa recibirá 3 argumentos: ruta al fichero original, ruta al fichero con la tabla de substituciones y ruta al fichero de salida.
- Las entradas de la TS se aplican en orden secuencial, por lo que se pueden obtener resultados diferentes con ordenaciones diferentes.

En las siguientes versiones, las funcionalidades que se desean ir incorporando son:

- **Entorno gráfico**: con una caja para escribir o pegar texto, una opción para cargar una tabla de substituciones y una caja de salida de texto.
- Opciones para **mantener la tabla de substituciones** (GUI).
- Posibilidad de utilizar **patrones de cadenas**.
- Diferentes métodos de proceso para cada regla (para cada entrada de la TS)

Tabla de substitución

Este componente es la razón de ser de este proyecto y el que realmente lo diferencia de otros proyectos similares. Cada tabla contendrá una información general (como la descripción) y una colección de reglas de substitución. Los atributos de los datos generales son:

Descripción	Título descriptivo de la tabla. Es el que se mostraría si la cargáramos a través de un desplegable.
Observaciones	Para incluir información como autor, fecha de creación, etc.
Versión	Del esquema empleado. En previsión de que en un futuro se pueda modificar la estructura de esta tabla, aquí indicamos qué estructura contiene.
Cadena inicial	Para añadir al principio.
Cadena final	Para añadir al final.
Condición	...para añadir las cadenas anteriores: 1-Siempre. 2-Solo si se ha producido alguna substitución

Los atributos de las entradas son los siguientes:

Etiqueta Origen Inicial	Cadena que se busca en el texto de entrada. Si no aparece, no se hace nada (se salta a la siguiente entrada de la tabla).
Etiqueta Origen Final	Una vez encontrada la cadena de entrada, se busca esta etiqueta final.
Etiqueta Destino Inicial	Es la que substituirá a la etiqueta origen inicial.
Etiqueta Destino Final	Es la que substituirá a la etiqueta origen final.
¿Eliminar?	Para eliminar todo el contenido entre las Etiquetas Origen Inicial y Final
Tipo Proceso	En la versión inicial solo hay uno, pero en un futuro se podrán definir diferentes maneras de realizar el proceso de búsqueda y substitución.
Observaciones	Comentarios para documentar esta entrada (no se tiene en cuenta).