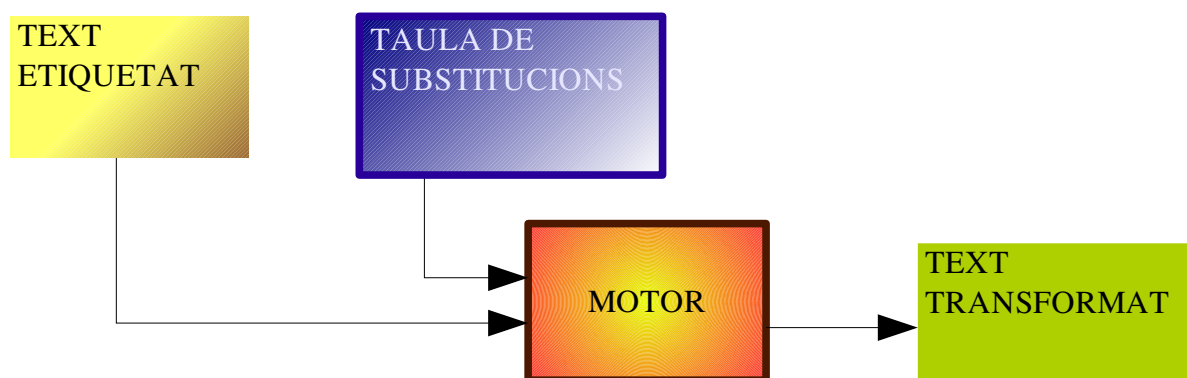


## TDTxE: Motor Transformador de documents (etiquetats)

### Resum

Aquest projecte neix a partir d'una necessitat concreta per convertir text amb etiquetes html4+css (amb estils) cap a html 3.2, degut a un problema d'incompatibilitat amb una aplicació que l'ha de manipular.

Aquesta necessitat es veu com una oportunitat per construir una aplicació més genèrica, que pugui transformar les etiquetes de qualsevol document de text a partir d'unes taules de transformació que els propis usuaris puguin elaborar. L'esquema de l'aplicació seria el següent:



Els dos components principals són:

-**Motor de procés:** que a partir d'un text d'entrada li aplica una sèrie de substitucions.

-**Taula de substitucions:** conté la relació de totes les substitucions que es volen aplicar amb tota la informació addicional que calgui. Aquestes taules han de poder ser fàcilment manipulables per l'usuari per tal de poder construir diferents taules de *conversió* i poder escollir en cada moment quina es vol utilitzar amb un determinat text.

També es necessiten 2 modes de funcionament:

-**Línia de comandaments:** rep com a paràmetres els documents d'entrada, sortida i taula de substitucions. Pensat per poder fer operacions *batch* o cridats des d'altres processos.

-**Interfície gràfica:** Amb una caixa per escriure/enganxar text i altra on surt el text processat, un cop s'ha escollit també la taula de substitucions a aplicar. Per utilitzar-la en casos en que es vol fer per exemple una "transformació ràpida" a partir de fragments d'un document, text resident a una altra aplicació i en definitiva, casos en que no existeix pròpiament un fitxer a convertir.

Utilitzant les taules adequades i amb el mateix motor, es pot utilitzar també per netejar codi, eliminar comentaris...

## Motor de procés

Com és habitual als projectes de programari lliure, partirem d'una **primera versió bàsica** que podrà evolucionar per incorporar opcions, millores de rendiment, etc. també en funció del que la pròpia comunitat demani i vulgui aportar.

La versió inicial consistirà en un procés iteratiu: per cada entrada de la Taula de Substitucions (TS a partir d'ara), es processa tot el document entrat. Quan s'arriba al final de la taula de substitucions i un cop processada la seva darrera entrada, s'haurà acabat el procés i tindrem el fitxer transformat. Per tant, cal tenir present el següent (per la primera versió):

- Les substitucions són textuais és a dir, no s'apliquen patrons de cadenes.
- Només funcionarà en mode consola. El programa rebrà 3 arguments: ruta al fitxer original, ruta al fitxer amb la taula de substitucions i ruta al fitxer de sortida.
- Les entrades de la TS s'apliquen en ordre seqüencial, pel que es poden obtenir resultats diferents amb ordenacions diferents.

A les següents versions, les funcionalitats que es volen anar incorporant són:

- **Entorn gràfic**: amb una capça per escriure o enganxar text, una opció per carregar una taula de substitucions i una capça de sortida de text.
- Opcions per **mantenir la taula de substitucions** (GUI).
- Possibilitat d'utilitzar **patrons de cadenes**.
- Diferents mètodes de procés per cada regla (per cada entrada de la TS)

## Taula de substitució

Aquest component és la raó de ser d'aquest projecte i el que realment el diferencia d'altres projectes similars. Cada taula contindrà una informació general (com la descripció) i una col·lecció de regles de substitució. Els atributs de les dades generals són:

<b>Descripció</b>	Títol descriptiu de la taula. És el que es mostraria si la carreguéssim a través d'un desplegable.
<b>Observacions</b>	Per incloure informació com l'autor, data de creació, etc.
<b>Versió</b>	De l'esquema emprat. En previsió de que en un futur es pugui modificar l'estructura d'aquesta taula, aquí indiquem quina estructura conté.
<b>Cadena inicial</b>	Per afegir al començament.
<b>Cadena final</b>	Per afegir al final.
<b>Condicó</b>	...per afegir les cadenes anteriors: 1-Sempre. 2-Només si s'ha produït alguna substitució.

Els atributs de les entrades són els següents:

<b>Etiqueta Origen Inicial</b>	Cadena que es busca al text d'entrada. Si no apareix, no es fa res (es salta cap a la següent entrada de la taula).
<b>Etiqueta Origen Final</b>	Un cop trobada la cadena d'entrada, es busca per aquesta etiqueta final.
<b>Etiqueta Destí Inicial</b>	És la que substituirà l'etiqueta origen inicial.
<b>Etiqueta Destí Final</b>	És la que substituirà l'etiqueta origen final.
<b>Eliminar?</b>	Per eliminar tot el contingut entre les Etiquetes Origen Inicial i Final
<b>Tipus Procés</b>	En la versió inicial només n'hi ha un, però en un futur es podran definir diferents maneres de realitzar el procés de cerca i substitució.
<b>Observacions</b>	Comentaris per documentar aquesta entrada (no es té en compte).