

XML - Editeur (Projet Caderige)

Documentation préliminaire de la Version 5.22

Objectifs de l'éditeur

Cet éditeur permet d'annoter interactivement et de manière graphique tout ou partie d'un document textuel. L'annotation consiste à étiqueter des zones de textes à l'aide de balises qui sont spécifiées par une DTD externe à l'éditeur. Le langage d'annotation utilisé ici est décrit dans le document "Format d'annotation pour CADERIGE". Une fois annoté, le texte peut être exporté au format XML et être utilisé dans d'autres outils ...

Lancement de l'application

Il faut tout d'abord décompacter l'archive ZIP. On obtient un dossier nommé "XmlJava5.21" qui contient l'application et des fichiers d'exemple. Pour pouvoir utiliser le système on doit disposer sur la station de travail d'une machine JAVA récente (\geq version 1.3).

Sous UNIX (ou Linux) il faut ouvrir un terminal de commande (Xterm) et se placer dans le dossier "XmlJava5.21/bin/". On lance alors le système avec la commande suivante :

```
java -classpath SgmlJava.jar MainFrame
```

Sous d'autres OS (OSX notamment, Windows ?) il est possible de lancer l'application en double-cliquant simplement sur l'exécutable *SgmlJava.jar*

Après quelques instants la fenêtre de l'éditeur apparaît ...

Organisation de l'interface

La fenêtre est organisée en 4 parties contenant respectivement :

- * Le texte en cours d'annotation (haut à gauche)
- * Les étiquettes XML qui ont été introduites dans le texte (bas à droite)
- * La valeur des attributs de l'étiquette courante (haut à droite)
- * La liste des étiquettes que l'on peut utiliser (bas à droite)

Ces zones peuvent être facilement redimensionnées en faisant glisser les séparations. Initialement les quatre zones sont vides, pour utiliser l'interface il faut :

- 1) Charger une DTD (par défaut c'est "Caderige.DTD" qui est chargé au lancement)
- 2) Charger un nouveau texte ou un texte déjà partiellement annoté.

Chargement des fichiers

Les commandes de chargement de fichiers se trouvent dans le menu FILE et sont également accessibles sous la forme d'icônes en haut de l'interface, sous la barre de menus.



: Chargement d'une DTD (par exemple "Caderige.DTD")



: Chargement d'un fichier text non encore annoté (par exemple "Short-text.txt")

Gestions des styles

Lorsqu'une DTD est chargée la liste des étiquettes (encore appelées "balises") qui sont accessibles apparait dans l'interface. Il est possible d'associer à chacune d'elle un "style d'affichage" en cliquant sur l'icone qui est situé à droite de chacune d'elles. Cette notion de styles est identique à celle que l'on trouve dans les traitements de textes ...

Par exemple, on peut modifier le style de la balise "Interaction". Après avoir cliqué sur l'icone associée à cette balise, une fenêtre apparait qui permet de sélectionner la police, la taille, l'aspect et les couleur (fond et forme) de la zone qui correspond à cette étiquette.

L'ensembles des styles peut être sauvé via les commandes du menu BALISE :



: Sauvegarde des styles courants associée aux balises



: Chargement d'un ensemble de styles préalablement définis

Principes d'annotations d'un texte

Le principe d'utilisation de l'éditeur repose sur la sélection à l'aide de la souris de la zone de texte (ou le paragraphe) que l'on souhaite annoter puis sur l'application d'une balise à cette zone. Pour se faire il suffit de cliquer sur l'étiquette correspondante dans la liste. La zone ainsi distinguée apparait alors dans le "style d'affichage" qui a été associé à la balise.

Par ailleurs, en fonction de la DTD utilisée, un certain nombre d'attributs servant à préciser la signification de l'annotation peuvent être associés à une balise. Cas échéant, ils apparaissent dans la partie supérieure droite de la fenêtre et peuvent être modifiés par l'utilisateur. Par exemple quatre attributs sont associés à la balise "interaction" : ils permettent de préciser la catégorie, le nom, le degré de certitude et le type de l'interaction décrite dans le texte.

A tout instant la position du curseur détermine l'emplacement de la balise courante, dès lors pour examiner une annotation il suffit de placer le curseur dans celle-ci. Dans cette version de l'éditeur, la zone "active" est affiché en caractères gras.

Si la DTD le permet, les annotations peuvent être imbriquée les une dans les autres. Ainsi, on peut (et on doit même :-) mettre des balises "An" (désignant l'agent d'une interaction) dans une zone déjà annotée comme "interaction".

Pour supprimer une annotation il suffit de placer le curseur dans celle-ci puis de cliquer sur le bouton "Delete current tag" pour effacer l'étiquette courante ou sur le bouton "Delete tag group" pour supprimer *l'ensemble* des étiquettes de cette zone c'est à dire l'étiquette courante et toutes celles qu'elle contenait éventuellement (le sous-arbre XML)

La zone inférieur droite de la fenêtre montre à tout instant les balises XML qui sont insérée dans le texte. Si l'on est pas intéressé par cette information on peut réduire cette partie jusqu'à la faire disparaître. Il faut noter que la navigation entre les deux fenêtres de gauche est

"synchronisée", en d'autre termes, lorsqu'on déplace le curseur dans l'une des zones l'autre est automatiquement décalée (si nécessaire) de manière à visualiser la même information.

Validation et sauvegarde

L'éditeur laisse l'utilisateur libre d'annoter son texte dans l'ordre ou il le souhaite. A tout instant il est cependant possible de vérifier que les annotations effectuées sont cohérentes avec la DTD en demandant une validation. Deux options sont disponibles :



: Validation partielle (les parties annotées sont correctes vis à vis de la DTD)



: Validation stricte (l'ensemble du document est conforme à la DTD)

A tout instant l'utilisateur peut sauver le travail en cours (qu'il soit cohérent ou non avec la DTD) et le reprendre ultérieurement, cela s'effectue avec ces deux commandes :



: Sauvegarde du document courant



: Chargement d'un document préalablement sauvé

Une fois qu'un document est validé, on peut enfin l'exporter au format XML (menu FILE). Il faut noter que ce format n'étant pas pour l'instant lisible par l'éditeur, il s'agit bien d'une *exportation* et non d'une *sauvegarde* du fichier !

Contact

Ce système est développé par le projet CADERIGE (<http://caderige.imag.fr/>) dans le cadre du programme Bioinformatique. Pour toutes questions contacter : <gilles.bisson@imag.fr>