

XML - Editeur (Projet Caderige ☐ <http://caderige.imag.fr/>)

Manuel d'utilisation (Editeur version 9.2 du 11/07/03)

Contact ☐ gilles.bisson@imag.fr

Objectifs de l'éditeur

Cet éditeur permet d'annoter interactivement tout ou partie d'un document à l'aide de balises XML décrites dans une DTD. Voici quelques sites présentant ces technologies :

- Présentation du XML : <http://www.brics.dk/~amoeller/XML/overview.html>
- Ecriture d'une DTD : http://www.w3schools.com/dtd/dtd_intro.asp

Contrairement à ce qui est couramment proposé dans ce type d'éditeur, l'utilisateur n'a pas à structurer directement le texte sous la forme d'un arbre XML. Il peut introduire les balises une à une dans l'ordre où il le souhaite en restant dans une optique voisine de celle d'un « traitement de texte ». Les zones de textes balisés sont visualisées à l'aide feuilles styles qui librement éditables par l'utilisateur. Le logiciel étant écrite en JAVA (version ≥ 1.3), il est exécutable sur un large nombre de plateformes : Linux, MacOS X et Windows.

Installation de l'application

Il faut décompresser l'archive ZIP. On obtient un dossier nommé "Editeur-XML" qui contient l'application et des fichiers exemples organisés de la manière suivante ☐

| | |
|--------------------|--|
| • XmlJava.bat | <i>Script de lancement</i> |
| • SgmlJava.jar | <i>L'exécutable JAVA</i> |
| • Example.txt | <i>Un fichier exemple ...</i> |
| • Preferences/ | <i>Dossier de configuration du système</i> |
| - xxx.DTD | <i>DTD par défaut du projet XXX</i> |
| - Default.xmlStyle | <i>Style par défaut</i> |
| - ... | |
| • Icons/ | <i>Icônes utilisées dans le logiciel</i> |
| • Splash/ | <i>Ecran d'accueil</i> |
| • Bugreport/ | <i>Dossier contenant les « Bugs »</i> |

Afin de pouvoir utiliser le système on doit disposer sur la station de travail d'une machine JAVA récente (version 1.3 minimum). Pour lancer l'éditeur sous UNIX (ou Linux) il faut ouvrir un terminal de commande (Xterm), se placer dans le dossier « XML-EDIT-x.x » puis exécuter la commande suivante :

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

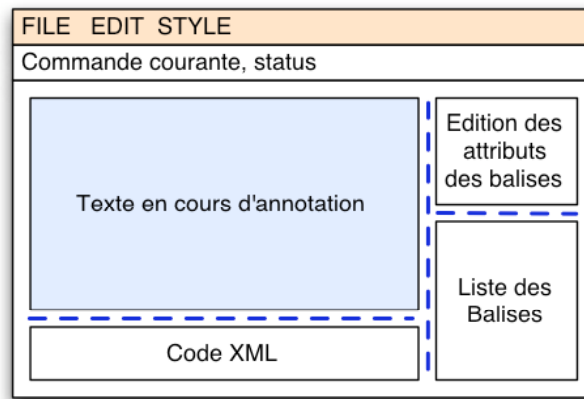
Sous d'autres OS (OS X notamment, Windows) il est possible de lancer l'application en double-cliquant simplement sur l'exécutable *SgmlJava.jar*. Après quelques instants, la fenêtre principale de l'éditeur apparaît ...

Organisation de l'interface

Outre la barre de menu et la ligne d'outils contenant quelques commandes usuelles, l'indicateur de position du curseur et les commandes de validation, la fenêtre est organisée en

quatre zones qui peuvent être redimensionnées en faisant glisser les barres de séparations

- **TEXT** : Le document initial en cours d'annotation (haut à gauche)
- **XML** : Les étiquettes XML qui ont été introduites dans le texte (bas à gauche)
- **ATT** : La valeur des attributs de la balise courante (haut à droite)
- **BAL** : La liste des balises utilisables (bas à droite) dans le contexte courant



Lorsqu'il y a une DTD et un fichier de styles par défaut dans le dossier préférence, la liste des balises est automatiquement initialisée lors du lancement de l'éditeur. Ensuite, pour pouvoir commencer à travailler il faut

- **Charger un nouveau texte ASCII «Vierge»**
- **Reprendre un texte déjà partiellement annoté.**

Dans la version actuelle de l'éditeur, pour des raisons de performances, il est préférable de ne travailler que sur des documents « courts » contenant moins de 3000 mots. Pour commencer à utiliser l'éditeur on peut charger le fichier « example.txt » à l'aide de la commande « Load text document » qui se trouve dans le menu « File ».

Principes d'annotation d'un texte

Le principe d'utilisation de l'éditeur repose sur deux étapes 1) la sélection à l'aide de la souris de la partie du document (mot, paragraphe, ...) que l'on souhaite annoter puis 2) sur l'application d'une balise à cette partie. Pour se faire, il suffit de cliquer sur l'étiquette voulue dans la liste BAL des balises. La zone ainsi distinguée apparaît alors dans le « Style d'affichage » qui est associé à la balise (ce style étant évidemment modifiable par l'utilisateur) et une nouvelle balise apparaît dans la zone XML en bas de l'éditeur.

A un instant donné, en fonction de la position du curseur, seules les balises de la DTD qui sont *immédiatement* plus spécifiques que la balise courante peuvent être insérées, les autres sont colorées en gris et ne peuvent être introduites. L'avantage de ce fonctionnement¹ est qu'il permet de limiter les erreurs de balisage et de guider l'utilisateur dans l'annotation.







Selon la DTD utilisée, chaque balise peut avoir un certain nombre d'attributs qui servent à préciser sa signification. Ces attributs et leurs valeurs apparaissent dans la partie ATT de

¹ Notons que ce comportement est spécifique de cette version de l'éditeur, dans les prochaines versions on pourra si on le désire insérer toutes les balises plus spécifiques indépendamment du leur niveau dans le sous arbre de la DTD, ainsi l'utilisateur sera plus libre de faire l'annotation dans l'ordre qu'il souhaite.

l'éditeur et peuvent être modifiés par l'utilisateur. Si les valeurs possibles appartiennent à un type énuméré (liste), ces valeurs sont sélectionnables à partir d'un menu, dans le cas contraire une simple zone d'édition permet de saisir le texte associé à l'attribut.

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. Le nom et la liste des attributs de la balise courante s'affichent dans la partie ATT de l'éditeur.

La zone inférieure gauche de l'éditeur est découpée en trois onglets

-  XML tree  affiche les balises XML insérées dans le texte. Elle permet de contrôler si l'annotation que l'on est en train de faire est satisfaisante (notamment en ce qui concerne l'imbrication des balises). La navigation entre les deux fenêtres TXT et XML est «Synchronisée», en d'autres termes, lorsqu'on déplace le curseur dans l'une de ces zones l'autre est automatiquement décalée de manière à visualiser la même information.
-  Validation result  affiche les erreurs de validations (ligne concernée et message) lorsqu'on demande à l'éditeur de contrôler la correction du balisage qui a été effectué.
-  Tagged document  affiche après une validation le document XML dans son ensemble tel qu'il sera sauvé sur le disque via la commande «Save annotated document»

Modification d'un texte et de son balisage






Comme dans tout éditeur, l'utilisateur est libre d'insérer, copier/coller², supprimer du texte dans le document. 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).

Il peut toutefois arriver que plusieurs balises soient appliquées sur une même zone de texte. Dans ce cas pour indiquer à l'éditeur le niveau sur lequel on veut travailler, il suffit de cliquer dans la zone XML sur la balise concernée. L'éditeur fonctionne selon ce schéma

- 1^{er} click : sélectionne la zone balisée correspondante dans le texte,
- 2nd click : le curseur se positionne en début de cette zone,
- 3^{ème} click : le curseur se positionne à la fin de cette zone (puis on boucle début-fin).



Gestion des fichiers (Menu FILE)

Les commandes de chargement de fichiers se trouvent dans le menu FILE et sont (pour une partie d'entre elles au moins) directement accessibles sous la forme d'icônes en haut de l'interface, sous la barre de menus.

-  : Création d'un nouveau document "vide" à annoter
-  : Chargement d'un fichier «Texte» non encore annoté (par ex. "Exemple.txt")
-  : Chargement d'un fichier en cours d'annotation (XML)
-  : Sauvegarde d'un fichier en cours d'annotation (XML)
-  : Chargement d'une nouvelle DTD



² Dans cette version, seul le copier/coller de textes est implémenté. La prochaine version proposera un copier/coller de blocs annotés ce qui permettra de dupliquer des parties du document en conservant l'ensemble des balises XML.

Les fichiers annotés peuvent-être exportés dans un format (HTML ou RTF) en conservant l'aspect graphique (mais ils ne pourront pas être relus par l’éditeur sous cette forme).

 : Sauvegarde en fichier HTML
 : Sauvegarde en fichier XML




Validation du document




L'éditeur laisse l'utilisateur libre d'annoter les parties du document dans l'ordre qu'il désire. Il est cependant possible de vérifier que les annotations effectuées restent 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 parfaitement conforme à la DTD)







Il est important de souligner qu'à tout instant l'utilisateur peut sauver le travail en cours, qu'il soit cohérent ou non avec la DTD, puis le reprendre ultérieurement.

Gestions des styles d’affichage

A chaque balise de la DTD on peut être associé un « style d’affichage qui permet de l'identifier dans le texte annoté. Cette notion de styles est voisine à celle que l'on trouve dans les traitements de texte. Le menu style contient les deux commandes suivantes

 : Chargement d’une feuille de styles (fichier avec l’extension « .xmlStyle »)
 : Sauvegarde du style courant (pour faire un nouveau style par défaut)
 : Edition du style courant (pour en modifier l’aspect et le comportement)

Après avoir sélectionné la commande d’édition du style courant, une nouvelle boîte de dialogue apparaît qui est constituée de deux parties

-  gauche la liste des balises contenues dans la DTD en mémoire
- A droite un panneau avec deux onglets «GraphicStyle et «SemanticStyle

L’onglet «GraphicStyle

Cet onglet, actif par défaut, permet d’associer à la balise sélectionnée dans la partie gauche une police, une taille, une option d’affichage (gras, italique, souligné, ...) et des couleurs de caractère et de fond. Une fois la boîte de dialogue refermée les nouveaux styles sont appliqués dans le document annoté. Il faut noter que les modifications effectuées sur les styles de balises sont automatiquement sauvées avec le document annoté.

Si l’on veut sauver les modifications effectuées afin de construire un nouveau fichier de style par défaut il faut utiliser explicitement la commande sauvegarde du style.

L’onglet «**SemanticStyle**»

Au cours de l’annotation, l’ajout des balises XML visant à structurer³ le document (par exemple, pour indiquer un titre, les limites d’un paragraphe, ...) peut devenir rapidement fastidieux pour l’annotateur et de ce fait être assez systématiquement oublié. Afin de limiter ce problème l’éditeur propose un mécanisme de « *complétion automatique* » qui permet d’associer à chaque balise des directives (optionnelles) permettant d’insérer automatiquement la balise de niveau immédiatement supérieur dans la DTD lorsqu’elles sont absentes et de choisir à quels emplacements s’effectuent cette insertion.

Dans la boîte de dialogue, on trouve deux séries d’options imbriquées. La première, permettant de décider si l’absence de l’étiquette de niveau supérieur déclenche une action automatique ou non, puis, si c’est le cas, pour indiquer à quelles positions seront insérées les balises de début et fin manquantes. Il y a quatre options possibles dans l’éditeur :

- ☐ Ajout directement autour de la balise courante
- ☐ Ajout autour du paragraphe courant (délimité par les caractères <retour chariot>)
- ☐ Ajout au début et à la fin du document
- Ajout en se basant sur deux expressions régulières, décrivant, à partir du point d’insertion courant, le motif syntaxique à rechercher. L’insertion de la balise peut alors s’effectuer au choix avant ou après ces motifs.

Il est important de noter que cette gestion de l’insertion des balises manquantes est effectuée de manière *réursive* jusqu’à ce qu’on arrive à la racine de la DTD ou jusqu’à ce que l’on retombe sur une balise qui était effectivement prévue par la DTD. Les attributs des balises automatiquement insérés, prennent les valeurs par défaut.

Gestions des préférences de l’éditeur

Dans le menu «**Edit**», la commande «**Preferences**» permet de modifier le comportement par défaut de l’éditeur. La boîte de préférence est structurée selon 3 onglets :

L’onglet «**File**» (sélection des fichiers par défaut)

- ☐ Working Folder : Nom du dossier où il faut lire/sauver les documents
- ☐ Default DTD : Fichier contenant la DTD par défaut
- ☐ Default Style : Fichier contenant les Styles par défaut

L’onglet «**Font**» (sélection des polices par défaut)


- ☐ Plain text : Police utilisée dans la zone d’annotation
- ☐ Tag body : Police utilisée pour représenter les balises en XML
- ☐ Sample text : Police utilisée pour représenter le texte d’exemple en XML

L’onglet «**Options**» (paramètres divers du système)

- ☐ Ending balise : Affichage des balises de fin dans la zone XML (peu utile)
- ☐ Strict DTD tree : L’éditeur contrôle l’introduction des balises de manière à ce que seules les balises «**plus spécifiques**» que celle déjà

³ Par opposition aux balises qui ont pour objectif d’introduire une « sémantique » et fournir une interprétation des éléments présents dans le document, par exemple : désigner une entité biologique, un processus, ...

présente de la zone courante puisse être insérées. Lorsque cette option est active seules les balises «**immédiatement**» plus spécifiques (les enfants de la balise courante) restent sélectionnables. Ce mode permet de mieux guider l’utilisateur novice qui ne connaît encore pas la structure de la DTD.

-  Sample text length : Nombre de caractères de l’exemple à afficher (zone XML)
-  Tag length : Nombre de caractères de la balise à afficher (zone XML)

Contact

Ce système est développé par le projet CADERIGE (<http://caderige.imag.fr/>) qui est financé dans le cadre d’une ACI Bioinformatique. Pour toutes questions, remarques, propositions de modification de l’éditeur, il faut contacter : gilles.bisson@imag.fr

Par ailleurs, si vous constaté un problème ou anomalie quelconque dans l’éditeur, afin de faciliter sa correction, il faut nous transmettre le fichier sur lequel vous étiez en train de travailler au moment du problème ainsi que le fichier « log » correspondant (ils sont datés) qui se trouve dans le dossier Bugreport.