

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

Manuel d’utilisation (Editeur version 1.16 du 22/07/04)

Contact : gilles.bisson@imag.fr

Objectifs de l’éditeur

Cet éditeur permet d’annoter interactivement tout ou partie d’un document textuel 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 « surlignant » la partie du texte qui l’intéresse et en y appliquant la balise de son choix parmi celles disponibles. On reste ainsi dans une optique voisine de celle d’un « traitement de texte ». Les zones de textes balisés sont visualisées à l’aide d’une combinaison de polices et couleurs qui sont décrites dans des « feuilles de styles » librement configurables par l’utilisateur.

Configuration minimale

Le logiciel étant écrite en JAVA, il est exécutable sur un large nombre de plateformes : Linux, MacOS X et Windows. La machine virtuelle doit être récente : version $\geq 1.4.x$ et l’on doit disposer d’une quantité de mémoire raisonnable (≥ 256 Mo).

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 (les noms des fichiers de DTD et de Styles peuvent varier selon les distributions) :

- | | |
|---------------------------------|---|
| • XmlJava.bat | <i>Script de lancement</i> |
| • SgmlJava.jar | <i>L’exécutable JAVA</i> |
| • Example.txt | <i>Un fichier exemple ...</i> |
| • Preferences/
- Default.DTD | <i>Dossier de configuration du système
DTD par défaut</i> |
| - Default.Style | <i>Style par défaut</i> |
| - Xmljava.conf | <i>Configuration de l’éditeur</i> |
| - ... | |
| • Icons/ | <i>Icônes utilisées dans le logiciel</i> |
| • Splash/ | <i>Ecran d’accueil</i> |
| • Bugreport/ | <i>Dossier contenant les « logs »</i> |

Historique des versions

Nouveautés de la version 1.16

- Correction d’un problème de fuite mémoire introduit dans la version 1.13
- Amélioration de la rapidité de lancement de l’application
- Amélioration de la réactivité globale de l’éditeur
- Amélioration des différents modes d’affichage de la liste des balises
- Correction de divers bugs d’affichage

Nouveautés de la version 1.13

- Réorganisation dans la boîte de dialogue « Préférences » de l’onglet des options
- Amélioration de l’affichage de la liste des balises applicables avec la possibilité de ne faire apparaître dans la liste que les balises qui sont actives.
- Possibilité de contrôler la génération (ou non) des attributs dans le code XML
- Ajout d’un UNDO/REDO sur la dernière insertion de balise
- Amélioration de la robustesse de la lecture des fichiers, notamment lorsque les feuilles de styles ne correspondent pas totalement à la DTD en mémoire
- Améliorer de la gestion des fichiers et ajout d’un « Save_as »
- Nouvelles options dans la ligne de commande de l’éditeur pour nommer le fichier de départ (vide) et charger un fichier annoté: –name et –annot
- Options pour contrôler l’apparition de la boîte de dialogue de saisie des attributs
- Correction des bugs présents dans la version 1.08 ... Notamment ceux concernant le rechargement de fichiers annotés qui nécessitait le passage par un document vide et la mémorisation des options dans le « semantic style ».

Nouveautés de la version 1.08

- Ajout d’une boîte de dialogue pour saisir les attributs lors de l’insertion d’une balise
- Nouvelle option pour compléter automatiquement la sélection sur les mots
- L’éditeur ne génère plus de fichiers LOG (qui étaient inutiles)
- La fermeture de la fenêtre de l’éditeur affiche une boîte de confirmation

Nouveautés de la version 1.06

- Passage d’arguments en paramètres lors du lancement de l’éditeur
- Correction des bugs sur l’affichage des attributs de type REQUIRED
- Ajout de nouvelles options de configuration de l’éditeur
- Correction des bugs sur les styles et l’inversion de balises
- Complétion de la sélection aux mots les plus proches via la touche ALT

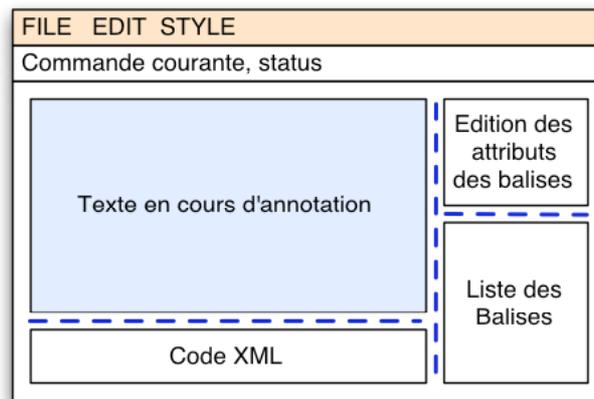
Nouveautés de la version 1.0 (versus 0.92)

- Homogénéisation de l’aspect de l’interface de l’éditeur
- Possibilité de personnalisation de la barre des outils
- Gestion simultanée de plusieurs styles en mémoire
- Nouveau format de sauvegarde des styles (compatibilité ascendante)
- Amélioration de la réactivité de l’éditeur sur les grands documents
- Fonction de recherche (Find)
- Correction de l’exportation en HTML

Organisation générale de l'interface

Outre la barre de menu et la ligne d'outils contenant les 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 :

- **TXT** : Le document initiale 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. Le nom des fichiers par défaut peut être modifié via l'option Préférence dans le menu Edit (cf. « gestion des préférences »). Ensuite, pour pouvoir commencer à travailler il faut :

- Charger un nouveau texte ASCII « vierge »
- OU
- Reprendre un texte déjà partiellement annoté.

Dans la version actuelle de l'éditeur (cette limitation sera levée dans les versions ultérieures), pour des raisons de performances, il est préférable de ne travailler que sur des documents relativement « courts » contenant moins de 10000 mots.

Pour commencer à utiliser l'éditeur on peut charger le fichier « exemple.txt » à l'aide de la commande « Load text document » qui se trouve dans le menu « File ».

Principes élémentaire de l’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 « format d'affichage » qui est associé à la balise et une nouvelle balise apparaît dans la zone XML en bas de l'éditeur. Le format d’affichage de chacune des balises est évidemment modifiable en éditant la « feuille de style » qui lui est associée (cf menu Style).

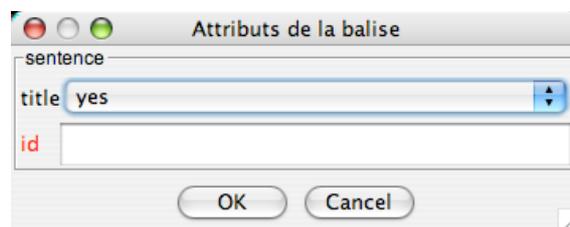
Pour faciliter la sélection, la zone courante est automatiquement ajustée sur la limite (espace ou ponctuation) des mots immédiatement à gauche et à droite. C’est le comportement par défaut dans l’éditeur, mais il est possible de le modifier via une option dans la boîte de préférences (cf description du menu « Edit ») et dans ce cas l’ajustement de la sélection ne s’effectuera que lorsque l’on appuie sur la touche AL.

Initialement : We isolated a temperature-sensitive sporulation defective mutant of the #SigA#
 + ALT : We isolated a temperature-sensitive sporulation defective mutant of the #SigA#

A un instant donné, en fonction de la position du curseur, seules les balises de la DTD qui sont *plus spécifiques* que la balise courante peuvent être insérées, les autres sont colorées en gris et ne peuvent être introduites (différentes configurations de présentations existent, voir la « gestion des préférences »). 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 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.

Lors de l’insertion d’une nouvelle balise une boîte de dialogue apparaît pour forcer l’utilisateur à saisir les différentes valeurs d’attributs (s’il y en a). Ce comportement est toutefois paramétrable via les préférences (cf. « gestion des préférences »). En cliquant sur le bouton « cancel » on annule l’insertion en cours. Les attribut de type REQUIRED (obligatoires) ont leur nom affiché en rouge (c’est le cas de l’attribut « Id » ci-dessous) :



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 affiche les balises XML insérées dans le texte. Elle permet de contrôler si l’annotation en cours est satisfaisante (notamment en ce qui concerne

¹ Ce comportement peut être modifié avec une option de configuration qui permet de choisir si l’on veut n’introduire que des balises *immédiatement* plus spécifiques ou non (cf description des préférences).

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.

Modification d'un texte et de son balisage

Comme dans tout éditeur, l'utilisateur est libre d'insérer, copier/coller, 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 (même position de début et de fin). Dans ce cas pour indiquer à l'éditeur le niveau sur lequel on veut effectivement 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. "Example.txt")
-  : Chargement d'un fichier en cours d'annotation
-  : Sauvegarde d'un fichier annoté (« Save »)
-  : Sauvegarde d'un fichier annoté en demandant le nom (« Save As ... »)
-  : Chargement d'une nouvelle DTD

Lors de la sauvegarde des documents annotés, l'éditeur construit deux fichiers : le premier (extension « .XML ») contient le texte annoté en XML et le second la description de la feuille de style courante² (avec une extension « .Style » qui dans cette version remplace le format précédemment utilisé « .XmlStyle »).

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). Par contre l'impression d'un document ne concerne que sa version XML.

-  : Sauvegarde en fichier HTML
-  : Sauvegarde en fichier XML

Lorsque l'on quitte l'éditeur une boîte de dialogue de confirmation demande à l'utilisateur s'il veut sauver le travail en cours.

² Une prochaine version de l'éditeur sauvera l'ensemble des styles chargés en mémoire

Edition et préférences (Menu EDIT)

Le menu Edit contient les commandes suivantes :

-  : Le Undo supprime la dernière balise qui a été insérée dans le document
-  : Le Redo rétablit la balise supprimée
-  : Coupe le texte sélectionné
-  : Copie le texte sélectionné
-  : Colle³ le texte sélectionné (sans les balises qu’il contient)
-  : Recherche d’une chaîne de caractères dans le document. La boîte de dialogue propose plusieurs options : sens de la recherche, recherche à partir du haut du document, prise en compte de la casse et recherche d’expressions régulières⁴.
-  : Positionne l’éditeur sur l’occurrence suivante dans la recherche
-  : Suppression de la balise courante (désignée par la position du curseur). Cette commande ne modifie pas les balises qui seraient incluses dans celle-ci. Par exemple si l’on supprime la balise ayant la feuille de style rouge :

Avant : `contrast, both |the E213G and E213A Spo0A variants showed`

Après : `contrast, both the E213G and E213A Spo0A variants showed`
-  : Suppression récursive de l’ensemble des balises dans la balise courante (désignée par la position du curseur). Cette commande affiche une boîte de confirmation car elle est susceptible de faire perdre la totalité des annotations effectuées dans un document.. Par exemple, dans le cas ci-dessus, on aurait :

Avant : `contrast, both |the E213G and E213A Spo0A variants showed`

Après : `contrast, both the E213G and E213A Spo0A variants showed`
-  : Gestion des préférences de l’éditeur (cf ci-dessous)

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 4 onglets :

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

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

³ Dans cette version, seul le copier/coller de textes est implémenté. Une prochaine version, l’éditeur proposera un copier/coller permettant de dupliquer des blocs en conservant l’annotation XML.

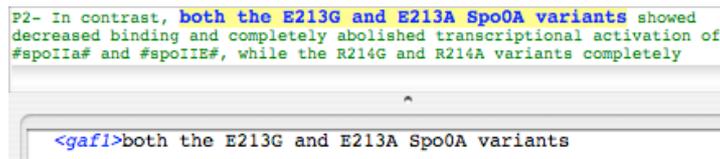
⁴ Les expressions régulières permettent d’exprimer des requêtes complexes. Pour plus de détail sur la syntaxe à utiliser on pourra se référer (entre autres) à ce site : <http://www.commentcamarche.net/php/phpreg.php3>

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

Le menu en haut de la fenêtre donne accès à trois options :

- **Plain text** : Police utilisée pour le texte non encore annoté
- **Tag body** : Police utilisée pour représenter les balises en XML
- **Sample text** : Police utilisée pour représenter le texte d’exemple en XML

Par exemple dans l’extrait ci-dessous le texte non encore annoté (« Plain text ») est en vert en taille 10 et dans la zone XML la balise (« Tag body ») apparaît en bleu italique et l’extrait associé (« Sample text ») est en noir :



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

La fenêtre est divisée en quatre zones différentes :

- *XML tree display*

- **Text size limit** : Taille maximale de l’échantillon de texte dans la zone XML.
- **Tag size limit** : Taille maximale de la balise et de ses attributs dans la zone XML.
- **Show ending balise in XML** :
Affichage des balises de fin dans la zone XML. Cette option est obsolete ...

- *Attribute options*

- **Use a dialog to enter attributes values** :
Permet de choisir si lorsqu’on introduit une nouvelle balise, une boîte de dialogue doit apparaître afin de forcer l’utilisateur à saisir la valeur des attributs. Cette boîte peut apparaître « toujours » dès qu’il y a un attribut (le comportement actuel), « jamais » (on utilise seulement la partie édition de la fenêtre) ou uniquement lorsqu’il y a au moins un attribut REQUIRED dans la balise.
- **Verify if REQUIRED attributes are set** :
Lorsque l’affichage de la boîte de saisie des valeurs d’attributs est actif, cette option permet de forcer l’utilisateur à toujours renseigner les attributs REQUIRED. La boîte ne se ferme alors que lorsque chacun d’eux a une valeur.
- **Allow to suppress all the attributes** :
Lorsque cette option est active une « check-box » apparaît à côté de chaque attribut dans les zones d’édition fenêtre principale et boîte de saisie. Elle permet de « désactiver » les attributs. Un attribut désactivés ne sera pas écrit dans le code XML à moins que ce soit un attribut REQUIRED. Cette option permet d’éviter de mettre dans le code XML des attributs qui ne sont pas valués.
- **Including REQUIRED attributes** :
Lorsque l’option précédente est activée on peut forcer l’éditeur à ne pas générer dans le code XML les attributs REQUIRED qui sont désactivés. Attention : avec cette option le code XML généré n’est plus conforme à la DTD.

- Tag options

- **Hide the unactive tags in the list :**
Si l’option est désactivée (défaut) les balises qui ne sont pas activables apparaissent en grisé. Si elle est activée, ces balises sont totalement « masquées » et disparaissent de l’affichage (laissant des emplacements vides).
- **Dynamically arrange the hidden tags :**
Si l’option précédente est activée, ce mode permet de réafficher la liste des balises de manière « compacte » en supprimant les emplacements vides. Ce mode est le plus agréable à utiliser lorsque l’on a une DTD complexe et longue.
- **Allows to insert all the most specific tag :**
Si l’option est désactivée (défaut) l’utilisateur ne peut sélectionner que des balises qui sont immédiatement plus spécifiques que la balise courante, sinon toutes les balise plus spécifiques sont acceptables. Le mode « strict » laisse moins de liberté dans l’ordre dans lequel l’annotation est faite, mais il est très utile pour guider ceux qui ne maîtrisent pas encore bien une DTD.

- Text options

- **Extend the selection to the borders of the word :**
Si cette option est active (défaut) la sélection d’une zone de texte est toujours ajustée sur les mots les plus proches ce qui simplifie la sélection des fragments de phrases. Dans le cas contraire, on peut forcer ce comportement en appuyant sur la touche ALT après que la sélection ait été effectuée.

• L’onglet « *Toolbar* » (Gestion de la barre des outils)

Le système permet de sélectionner les commandes qui apparaissent sous la forme d’icônes dans la barre des outils. Dans la partie gauche de la fenêtre on trouve la liste des commandes de l’éditeur et à droite on peut sélectionner si cette commande est visible ou nom et lui associer une position dans la barre.

La position d’une commande est déterminée par 2 valeurs : le n° de groupe (de 0 à 9) le groupe 0 étant celui le plus à gauche et sa position dans le groupe (0 à 9 également). La notion de groupe existe afin de pouvoir rassembler les commandes portant sur des opérations de type similaire (lecture/écriture de fichiers, éditions des balises, ...)

Gestions des styles d’affichage (Menu STYLE)

A chaque balise de la DTD on peut être associé un « style d’affichage » qui permet de l’identifier facilement dans le texte annoté. Cette notion de styles est similaire à celle que l’on trouve dans les traitements de texte (Word par exemple) ou avec les CSS dans les pages HTML. Le menu style contient les commandes suivantes :

-  : Chargement en mémoire d’une feuille de styles. Le chargement concerne des fichiers au format « xmlStyle » (ancien format) ou « .style » (nouveau format).
-  : Sauvegarde de la feuille de style courante (avec l’extension « .style »)
-  : Sauve la feuille courante comme nouveau fichier par défaut « default.style »

dans les préférences. Lors de la création d’un nouveau document ou lors du lancement de l’éditeur c’est ce nouveau style qui est normalement utilisé.



: Création d’une nouvelle feuille de styles



: Duplication d’une feuille de styles



: 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 :

- A 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 sauveés 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 structurer⁵ 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.

⁵ 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, ...

Menu HELP

Pour l'instant ce menu ne contient qu'une commande (ABOUT) permettant d'afficher le numéro de version et la date de création de l'éditeur.

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). Cette validation n'est utile que si le fichier a été modifié via un éditeur de texte, a priori sous l'éditeur d'annotation, la syntaxe est toujours correcte.

 : 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.

Contact

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

Par ailleurs, si vous constatez un problème ou anomalie quelconque dans l'éditeur, afin de faciliter sa correction, il faut nous transmettre les fichiers (document, DTD, style) sur lesquels vous étiez en train de travailler au moment du problème ainsi qu'une description précise du problème que vous avez rencontré.