



EXAMEN D'XML

Master — 2023-2024

Julien Lemaire

SOMMAIRE

Conditions de réalisation :	p.3
Exercice 1 : structuration en XML	p.4
Exercice 2 : encodage en XML TEI:P5	p.5
Exercice 3 : chaîne d'XMLisation	p.7

CONDITIONS DE RÉALISATION :

CRÉATION DES GROUPES

- Le travail peut être réalisé par groupe de 2 personnes (il peut aussi être réalisé seul, même si je préférerais vous voir travailler en groupe) [la charge de travail a été mesurée en considérant que deux personnes allaient faire le devoir].
- Vous déclarerez le groupe auquel vous appartenez dans un tableur hébergé sur Nextcloud à l'adresse suivante :

<https://nextcloud.univ-lille.fr/index.php/s/T4dNf8eFG7BAQqx>

DÉPÔT

- Vous déposerez le devoir sur Moodle dans un espace dédié au dépôt des devoirs situé dans la section « Examen ».
- Seule une des deux personnes est tenue de déposer le devoir. Ce sera la personne dont le nom de famille intervient le plus tôt dans l'ordre alphabétique.

DÉLAIS

- Vous bénéficiez de 4 semaines pour terminer le devoir et le déposer sur Moodle. La date de rendu est donc fixée au **5 décembre 2023 à 23 h 59** (15 décembre 2023 23 h 59 pour les tiers-temps).
- Si un des étudiants du groupe bénéficie d'un tiers-temps, le tiers-temps s'applique à tout le groupe.

DIVERS

- Vous veillerez à ajouter, en commentaire, au début de vos documents, les noms et prénoms des personnes qui ont contribué.
- Ne plagiez pas les camarades qui appartiennent à un groupe différent du vôtre (je serai intransigeant sur ce point).
- Rendez des documents propres et indentés avec des tabulations.
- Il est impératif que vous commenciez le devoir le plus tôt possible sous peine de vous retrouver dans une position plus que délicate au moment du rendu.

EXERCICE 1 : STRUCTURATION EN XML

Pour ce premier exercice, vous devez élaborer une structure sémantique capable d'encoder un extrait du dictionnaire *Le Robert junior illustré millésime 2017*. Vous pouvez vous aider du mode d'emploi conçu par les lexicographes pour vous aider à identifier les éléments, mais en veillant à utiliser une terminologie plus professionnelle que celle utilisée.

- Identifier les éléments structurels adéquats pour l'encodage d'un dictionnaire.
- Rédiger une DTD qui permette de valider le document XML.
- Rendre un document XML bien formé.
- Rendre un document XML valide conformément à votre DTD.
- Vous réaliserez également une représentation de votre structure sous la forme d'un arbre ou d'un tableau. L'essentiel sera de pouvoir identifier les relations hiérarchiques qu'entretiennent les éléments.

Fichiers fournis :

- /Exercice-1
 - ↳ EX1_LRJI-2017_sv-boulier_p118.pdf
 - ↳ EX1_LRJI-2017_mode-d-emploi_1.pdf
 - ↳ EX1_LRJI-2017_mode-d-emploi_2.pdf

Fichiers à rendre :

Les documents sont à ranger dans un répertoire (dossier) nommé « **Exercice-1** ». Le répertoire « Exercice-1 » sera lui-même rangé dans un répertoire nommé « **XML_examen__nom1-prenom1_&_nom2-prenom2** ».

- /Exercice-1
 - ↳ EX1_nom1-prenom1_&_nom2-prenom2.xml
 - ↳ EX1_nom1-prenom1_&_nom2-prenom2.dtd
 - ↳ EX1_arbre-structurel_nom1-prenom1_&_nom2-prenom2.pdf

EXERCICE 2 : ENCODAGE EN XML TEI:P5

Encoder en XML TEI un extrait de la pièce de théâtre intitulée *Vautrin*, par Honoré de Balzac, éditée par Delloye et Tresse et parue en 1840. Vous rendrez un document XML bien formé et valide, respectueux des recommandations de la TEI:P5. Cet encodage XML veut rendre compte de la version fac-similaire du présent texte.

CONSIGNES GÉNÉRALES

- Identifier les éléments de la TEI qui correspondent à notre volonté de structuration sémantique.
- Encoder le texte avec les balises appropriées.
- Obtenir un document XML bien formé.
- Obtenir un document XML valide, conformément au schéma de la TEI:P5.
- Rédiger une DTD qui permette de valider votre document (vous limiterez l'utilisation de déclaration en contenu mixte au maximum) [cette DTD sera, en quelque sorte, une version réduite du schéma de la TEI:P5].
- Respecter les conventions typographiques en vigueur pour les espaces typographiques.

CONSIGNES DE BALISAGE

Pour le <teiHeader> :

- Ajouter, dans l'élément approprié, la notice bibliographique du document source.
- Le chercheur principal pour ce projet est : Julien Lemaire.
- Vous êtes les éditrices/éditeurs de ce projet.
- L'éditeur est : l'Université de Lille.
- Vous êtes les responsables de l'édition XML/TEI.
- Julien Lemaire est le responsable de l'informatique éditorial.
- Julien Lemaire est le responsable du traitement des images.
- Cette édition est libre de droits.
- Préciser que la langue utilisée est le français.

Pour le texte :

- Utiliser des attributs pour numérotter les actes et les scènes.
- Utiliser des attributs avec des identifiants, afin de relier chaque interlocuteur à sa mention dans la distribution.
- Il faudra utiliser un attribut spécifique pour marquer les didascalies alignées à droite.

— En considérant que l'objectif finale de notre travail d'XMLisation est une édition HTML consultable en ligne. Nous souhaiterions pouvoir afficher chaque page du fac-similé en vis-à-vis de la réédition HTML. Pour cela, nous devons indiquer dans notre document XML, les changements de pages par rapport à la version source disponible sur Gallica. Vous utiliserez un élément <pb> pour *page beginning* avec un attribut @n pour le numéro de page, un attribut @xml:id pour identifier la page de manière unique, et enfin, avec un attribut @facs contenant le chemin d'accès au scan de la version fac-similaire. Les numéros de page sont indiqués entre crochets dans les versions ODT et PDF du texte. Exemple : pour la page numéro 5, nous obtiendrions un élément de la forme suivante à l'endroit où démarre la cinquième page de l'ouvrage :
<pb n="5" xml:id="p5" facs="scans/vautrin-page-5.jpg"/>

Fichiers fournis :

- EX2_balzac-vautrin-1840-delloye&tresse_brut.txt
- EX2_balzac-vautrin-1840-delloye&tresse_brut.odt
- EX2_balzac-vautrin-1840-delloye&tresse.pdf
- /scans
 - └─ vautrin-page-x.jpg

Fichiers à rendre :

Les documents devront être rangés dans un répertoire intitulé « Exercice-2 ».

- /Exercice-2
 - └─ EX2_nom1-prenom1_&_nom2-prenom2.xml
 - └─ EX2_nom1-prenom1_&_nom2-prenom2.dtd

CONSEIL :

Un chapitre dédié à l'encodage en XML TEI de textes dramatiques est disponible sur le site internet de la TEI à l'adresse suivante :
<https://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html>

Utiliser le document ODT avec des styles pourrait vous faire gagner du temps. On peut appliquer des styles avec l'outil de recherche et de remplacement de Libre Office. Le stylage des interlocuteurs pourrait se faire relativement rapidement de cette manière.

EXERCICE 3 : CHAÎNE D'XMLISATION

Encoder en XML le Corpus didactique des enseignements de stylage, XML et XSLT des masters de Lettres et de Sciences du langage de l'Université de Lille. Vous utiliserez la DTD de littérature générale fournie pour valider votre document. Vous enrichirez la DTD afin de pouvoir encoder le plus précisément possible les chapitres IV, V et la bibliographie.

ÉTAPE 1 : STYLAGE

- Réaliser le stylage des parties non-stylées (chapitre IV et V, ainsi que la bibliographie).
- Créer des styles de paragraphes et de caractères qui permettront de récupérer un document XML le plus fidèle possible à la DTD de littérature générale.
- Vous ajouterez le préfixe « UL_ » devant les noms de styles que vous aurez créés.

ÉTAPE 2 : XMLISATION

- Récupérer le document « content.xml » et procéder à son nettoyage à l'aide des feuilles de styles XSL fournies.
- Obtenir un document XML bien formé.
- Assurer la validation de votre document XML avec la DTD de littérature générale fournie.
- Enrichir la DTD de littérature générale afin de valider le document
- Insérer les espaces typographiques adéquates devant les ponctuations doubles.

ÉTAPE 3 : CONVERSION TEI:P5

- Réaliser une version TEI:P5 de votre document XML.
- Pour le chapitre IV, utiliser des éléments de structuration de textes par défaut. En d'autres termes, n'utilisez pas des éléments issus des recommandations TEI pour l'encodage de dictionnaires.
- Utiliser les éléments adéquats pour l'encodage des textes dramatiques lorsque cela sera nécessaire.

Fichiers fournis :

Les documents fournis sont rangés dans le répertoire intitulé « Exercice-3 » :

- /Exercice-3
 - └ EX3_corpus_brut.odt
 - └ EX3_corpus.pdf
 - └ litterature_generale_V6.dtd
 - └ /XSL
 - └ ODTversXML=Audrain+Gasiglia_ODT2017-02.xslt
 - └ XMLversXML=suppr_Ti.xsl

Fichiers à rendre :

Les documents devront être rangés dans un répertoire intitulé « Exercice-3 ».

- /Exercice-3
 - └ EX3_nom1-prenom1_&_nom2-prenom2.odt
 - └ EX3_LG_nom1-prenom1_&_nom2-prenom2.xml
 - └ EX3_LG_nom1-prenom1_&_nom2-prenom2.dtd
 - └ EX3_TEI_nom1-prenom1_&_nom2-prenom2.xml

CONSEILS :

Utiliser l'outil de recherche et de remplacement avec des expressions régulières pour les tâches suivantes :

- Corrections récurrentes ;
- Traitement des espaces typographiques si besoin ;
- Conversion en XML-TEI:P5 (remplacement de certaines balises récurrentes) .

Un tutoriel pour l'utilisation de l'outil de recherche et de remplacement de l'éditeur <oXygen/> est disponible sur la page Moodle du cours.