

INTRODUCTION À LA TEI : ALLER PLUS LOIN

Floriane Chiffolleau

Ingénieure Recherche et Développement (Inria)

13-14 avril 2021, URFIST de Rennes

Repository GitHub : [FloChiff/Introduction_TEI](https://github.com/FloChiff/Introduction_TEI)

TRANSFORMER SON CORPUS (XSLT)

[ALLER PLUS LOIN](#)

TRANSFORMER SON CORPUS (XSLT)

- XSLT (eXtensible Stylesheet Language Transformation) : langage de programmation, et plus précisément langage de transformation, qui consiste à transformer un document XML en un autre format
- XPath : langage de requête XML qui permet de naviguer au sein même de l'arbre
- Exemples de format de sortie :
 - XML : d'un fichier XML à un autre, par exemple pour modifier certains éléments de l'arbre pour le mettre à jour
 - HTML : d'un fichier XML à un fichier HTML, pour créer des pages web qui affichent le contenu de l'arbre
 - LaTeX : d'un fichier XML à un document LaTeX, pour permettre une sortie avec une mise en page maîtrisée, pour être ensuite possiblement transformé en PDF

TRANSFORMER SON CORPUS (XSLT)

- Il existe de multiples règles de transformation en XSLT, qui permettent d'arriver à divers résultats en fonction de ce qui est requêté :
 - Requêter au sein de l'arbre :
 - Se situer dans l'arbre XML : `<xsl:template>`
 - Appliquer les règles de transformations définies pour un élément : `<xsl:apply-templates>`
 - Appliquer les règles de transformations définies pour un cas donné d'une partie de l'arbre : `<xsl:call-templates>`
 - Faire apparaître du contenu :
 - Chercher le contenu d'une partie de l'arbre XML, balises et textes : `<xsl:copy-of>`
 - Chercher le texte contenu dans les balises appelés : `<xsl:value-of>`
 - Appel d'une boucle pour itérer sur les éléments, les uns après les autres : `<xsl:for-each>`
 - Affichage d'un élément sous certaines conditions : `<xsl:if>/<xsl:choose>/<xsl:when>/<xsl:otherwise>`
 - Création d'un mode pour faire apparaître plusieurs fois la même partie de l'arbre XML mais selon divers critères : `@mode`

TRANSFORMER SON CORPUS (TEI PUBLISHER)

[ALLER PLUS LOIN](#)

TRANSFORMER SON CORPUS (TEI PUBLISHER)

- TEI Publisher → outil de publication numérique de corpus
 - Dépend de *exist-db*, un système de gestion de base de données basé sur la technologie XML
 - Conforme aux TEI Guidelines et s'appuyant sur le *TEI processing model*, qui documente le modèle de traitement pour un élément donné
 - Transformation de fichiers XML pour un affichage web et une exportation sous divers formats (LaTeX, ePUB, PDF)
- Génération d'une application web
 - ODD : règles de transformation des éléments de l'arbre XML TEI (condition d'affichage, mode de transformation, rendu de sortie du contenu, etc.)
 - Template : modèle de présentation de données, mise en page de l'affichage, diverses propositions de templates faites par TEI Publisher

TRANSFORMER SON CORPUS (TEI PUBLISHER)

Quelques exemples de mise en pratique

- Applications développés entièrement à l'aide de TEI Publisher
 - [Van Gogh Letters](#)
 - [Shakespeare's plays](#)
 - [Early English Books](#)
 - [When the wall came down](#)
- Exploration de la [collection de démonstration](#) de TEI Publisher
- Importation de ses propres documents pour tester [l'aire de jeux](#)

STRUCTURER LA TEI (ODD)

ALLER PLUS LOIN

STRUCTURER LA TEI (ODD)

- Déclarer le jeu de balises de son encodage
- Deux techniques de création :
 - Avec un ou plusieurs fichiers XML du corpus
 - Avec [Roma](#), un outil de génération de customisation pour la TEI
- Quatre types de déclaration de classes TEI :
 - La classe n'est absolument pas utilisé → pas de déclaration et tous ses éléments sont exclus de l'arbre
 - La classe est un peu utilisé → déclaration et intégration des éléments utilisés avec @include
 - La classe est beaucoup utilisé → déclaration et exclusion des éléments non utilisés avec @except
 - La classe est entièrement utilisé → déclaration et aucune exclusion avec un @except vide

STRUCTURER LA TEI (ODD)

- Personnalisation de son schéma TEI :
 - Addition d'éléments : créer un élément n'appartenant pas au domaine TEI mais considéré nécessaire au corpus
 - Suppression d'éléments : enlever la possibilité d'utiliser certains éléments dans des situations données
 - Changement d'éléments : ajouter ou supprimer les éléments et/ou attributs qui peuvent être contenus dans un élément donné
 - Personnalisation d'attributs et de leurs valeurs : modification des règles d'apparition des attributs ou choix de valeurs définis pour un attribut donné

STRUCTURER LA TEI (ODD)

- Exemple de personnalisation d'attribut :
 - @rend → attribut utilisé pour de multiples éléments
 - Transcription avec beaucoup de mots supprimés (<deletion>) de manière toujours récurrente (rayé, recouvert ou effacé)
 - Déclaration du changement :

```
<elementSpec ident="del" mode="change">  
  <attList>  
    <attDef ident="rend" mode="change">  
      <valList type="semi" mode="replace">  
        <valItem ident="overwritten"/>  
        <valItem ident="strikethrough"/>  
        <valItem ident="erasure"/>  
        <valItem ident="none"/>  
      </valList>  
    </attDef>  
  </attList>  
</elementSpec>
```

DOCUMENTER LA TEI (ODD)

ALLER PLUS LOIN

DOCUMENTER LA TEI (ODD)

- Expliquer les choix faits, présenter ses balises, hiérarchie de l'arbre XML
- Deux types de documentation :
 - Documentation partielle, présentation des balises dont l'utilisation peut différer par rapport à ce que définit les TEI guidelines
 - Documentation exhaustive, présentation de l'arbre de A à Z, avec description de l'utilisation des éléments/attributs dans l'arbre

DOCUMENTER LA TEI (ODD)

- Documentation écrite avec les balises traditionnelles pour l'écriture en prose dans le <body> :
 - <div> : pour la hiérarchie des parties
 - <head> : pour les titres de parties
 - <p> : pour le contenu de la présentation et des descriptions
- Balises spécifiques pour les composants de la structure TEI :
 - <gi> : pour les éléments
 - <att> : pour les attributs
 - <val> : pour la valeur des attributs
 - <egXML xmlns="<http://www.tei-c.org/ns/Examples>"> : pour les exemples tirés de l'arbre XML

DOCUMENTER LA TEI (ODD)

Exemple d'une documentation tirée d'une ODD

```
<div>
  <head><gi>msContents</gi></head>
  <p>The only sub-element of <gi>msContents</gi> is <gi>msItem</gi>. We use it for three different purposes.</p>
  <p>First, the document date may be given inside <gi>docDate</gi>. This is the date originally written on the
    letter or manuscript by the author<note>If a date was written by a different person, for example by an
      archivist, it is not included in <gi>docDate</gi></note>. It may be identified as incorrect and thus
    differ from the <gi>origDate</gi> (origin date) in the <gi>history</gi> section, where the actual date
    (which can be a supposition) is declared. If no date is given on the manuscript, then <gi>docDate</gi> is
    left out and only <gi>origDate</gi> appears in the header. If <gi>docDate</gi> and <gi>origDate</gi> are
    the same, then both still appear.</p>
  <p>Second, <gi>msItem</gi> may be used to give a summary of the manuscript or other details written by a
    project member in one or more <gi>note</gi> elements with the <att>resp</att> attribute.</p>
  <p>Third, <gi>msItem</gi> may be used to give keywords for the manuscript. They are also included inside
    <gi>note</gi> elements—one for each keyword—with <att>resp</att> and
    <att>type</att>=<val>keyword</val>.</p>
  <p>Here is an example including all three cases : <egXML xmlns="http://www.tei-c.org/ns/Examples">
    <msContents>
      <msItem>
        <docDate when="1919-02-04"/>
        <note resp="#floriane.chiffolleau">This letter is joined with an annexe</note>
        <note type="keyword" resp="#floriane.chiffolleau">President Wilson</note>
      </msItem>
    </msContents>
  </egXML></p>
</div>
```

DOCUMENTATION

ALLER PLUS LOIN

RESSOURCES ÉLECTRONIQUES

- XSLT Tutorial : https://www.w3schools.com/xml/xsl_intro.asp
- Xpath Tutorial : https://www.w3schools.com/xml/xpath_intro.asp
- Chapitre 22 Documentation Elements : <https://tei-c.org/release/doc/tei-p5-doc/fr/html/TD.html>
- Documentation TEI Publisher : <https://teipublisher.com/exist/apps/tei-publisher/doc/documentation.xml?odd=docbook.odd>