

Introduction à la TEI : Pourquoi encoder ses corpus ?

Floriane Chiffolleau, Doctorante en Humanités Numériques, ALMAAnaCH/Le Mans Université

URFIST de Rennes, 24-25 novembre 2022

Repository GitHub : https://github.com/FloChiff/Introduction_TEI_2022

Historique de la création de la XML-TEI

Historique de la création de la XML–TEI

- ❑ Novembre 1987 : Création de la Text Encoding Initiative (TEI)
 - Pallier à la prolifération de systèmes divers et incompatibles pour la représentation numérique de textes
- ❑ Mai 1994 : Publication de la première version officielle des TEI Guidelines (P3)
 - Règles produites par le travail combiné de nombreux ateliers de réflexions et par les révisions et extensions faites à la version P1 produite en 1990
 - Utilisation du SGML pour rédiger en TEI
- ❑ Janvier 1999-Janvier 2001 : Création du consortium TEI
 - Avoir une organisation officielle qui maintient, développe et promeut la TEI
- ❑ Juin 2022 : Publication de la version P4 des TEI Guidelines
 - Introduction du format XML (créé en 1999), comme nouveau format de rédaction de la TEI
- ❑ Novembre 2007 : Publication de la dernière version des TEI Guidelines (P5)
 - Version révisée de P4 avec de nouveaux développements pour un certain nombre de domaines
 - Abandon du SGML pour un passage complet au format XML

Qu'est-ce que le format XML ?

XML ou eXtensible Markup Language (langage de balisage extensible)

- Langage pour le partage de données entre différents systèmes grâce à un format souple
- Langage pour le balisage de documents textuels, dont les chaînes de caractères sont délimitées par des étiquettes appelées *élément* qui sont encadrées par des chevrons (<, >)
- Langage extensible qui n'est pas prédéfini et qui permet la création de son propre jeu d'éléments, qui peut toujours être étendu

Comment ça fonctionne ?

La XML se structure sous forme d'arbre:

- Un élément racine
- Des balises imbriquées que l'on désignera comme des parents et des enfants
- Des balises, à la suite, désignées comme des frères/sœurs

Prologue de fichier : déclaration du fichier XML et de l'encodage des caractères

Exemple : `<?xml version="1.0" encoding="UTF-8"?>`

Comment ça fonctionne ?

Plusieurs règles importantes à respecter dans la rédaction d'un arbre XML

- Balise ouvrante/fermante : Une balise ouvrante devra toujours être suivie dans l'arbre par une balise fermante

Exemple : `<titre>Présentation de personnes</titre>`

- Sensibilité à la casse : l'écriture de la balise fermante devra être identique à celle de la balise ouvrante

Exemple correct : `<titre>Présentation de personnes</titre>`

Exemple incorrect : `<titre>Présentation de personnes</Titre>`

Comment ça fonctionne ?

Plusieurs règles importantes à respecter dans la rédaction d'un arbre XML

- Pas de chevauchement : il faut bien faire attention à l'imbrication des balises
- Attributs : les éléments peuvent être enrichis d'attributs et leurs valeurs

Exemple correct :

```
<personne><nom>Martin</nom></personne>
```

Exemple incorrect :

```
<personne><nom>Martin</personne></nom>
```

Exemple :

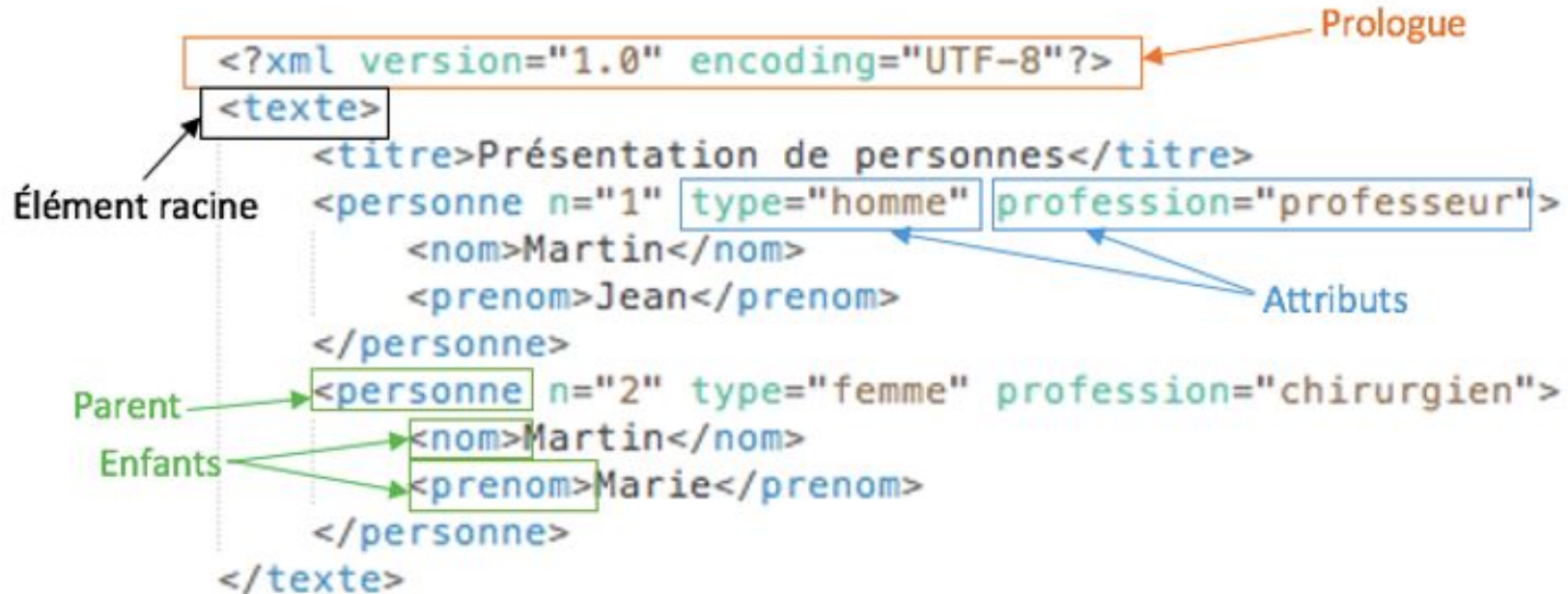
```
<personne n="1" type="homme" profession="professeur">  
  <nom>Martin</nom>  
  <prenom>Jean</prenom>  
</personne>
```

- Les valeurs des attributs doivent toujours être mises entre guillemets ou apostrophes

Exemple correct : `<personne type="homme">`

Exemple incorrect : `<personne type=homme>`

Comment ça fonctionne ?



Qu'est-ce que la TEI ?

Les TEI Guidelines

- ❑ Règles définies pour l'encodage de tous les textes sous leur format numérique et lisibles par machine
- ❑ Une liste longue, détaillée et en constante évolution
- ❑ De nombreuses balises pour encoder des corpus de type, langue et structure différentes (romans, pièces de théâtre, poème, lettres, rapport officiel, etc.)
- ❑ Page unique pour chaque balise et attribut, afin d'avoir toutes les informations sur la manière dont ils s'utilisent

Corps du texte

- + 1 [The TEI Infrastructure](#)
- + 2 [The TEI Header](#)
- + 3 [Elements Available in All TEI Documents](#)
- + 4 [Default Text Structure](#)
- + 5 [Characters, Glyphs, and Writing Modes](#)
- + 6 [Verse](#)
- + 7 [Performance Texts](#)
- + 8 [Transcriptions of Speech](#)
- + 9 [Dictionaries](#)
- + 10 [Manuscript Description](#)
- + 11 [Representation of Primary Sources](#)
- + 12 [Critical Apparatus](#)
- + 13 [Names, Dates, People, and Places](#)
- + 14 [Tables, Formulæ, Graphics, and Notated Music](#)
- + 15 [Language Corpora](#)
- + 16 [Linking, Segmentation, and Alignment](#)
- + 17 [Simple Analytic Mechanisms](#)
- + 18 [Feature Structures](#)
- + 19 [Graphs, Networks, and Trees](#)
- + 20 [Non-hierarchical Structures](#)
- + 21 [Certainty, Precision, and Responsibility](#)
- + 22 [Documentation Elements](#)
- + 23 [Using the TEI](#)

Un consortium collaboratif

jTEI : Journal de la TEI :

<https://journals.openedition.org/jtei/>

- ❑ Journal officiel du consortium TEI
- ❑ Articles sur l'état de l'art, des innovations en matière de TEI et des exploitations dans des projets
- ❑ Articles rigoureusement évalué avant publication sur la plateforme Open Edition

Conference/Annual meeting :

<https://members.tei-c.org/Events/meetings/>

- ❑ Partage et découverte des projets en cours sur la TEI
- ❑ Rencontre avec les autres membres de la communauté
- ❑ Suivi officiel du fonctionnement du bureau du consortium

Un consortium collaboratif

Special Interest Groups (SIGs) :

<https://tei-c.org/activities/sig/>

- ❑ Groupe de réflexion, d'échanges et de débats sur des sujets spécifiques liés à la TEI
- ❑ Aide au développement et à l'évolution de la TEI
- ❑ Trois exemples de SIGs :
 - ❑ Computer-Mediated Communication SIG
 - ❑ Correspondence SIG
 - ❑ East Asian/Japanese SIG

TEI-L mailing-list :

<https://tei-c.org/support/#tei-l>

- ❑ Ouverte à toute la communauté TEI, pour poser des questions ou y répondre
- ❑ Partage d'expertise ou d'expérience
- ❑ Archivage de tous les problèmes précédemment rencontrés

Documentation



Ouvrages

- ❑ HAROLD, Elliotte Rusty, MEANS, W. Scott, ENSARGUET, Philippe [et al.], *XML en concentré*, Paris, O'Reilly, 2005.
- ❑ BURNARD, Lou et BURGHART, Marjorie, *Qu'est-ce que la Text Encoding Initiative ?*, 2015 [en ligne : <http://books.openedition.org/oep/1237>].

Ressources électroniques

- ❑ TEI guidelines :
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/index.html>
- ❑ Journal de la Text Encoding Initiative :
<https://journals.openedition.org/jtei/>
- ❑ TEI Special Interest Groups (SIGs):
<https://tei-c.org/Activities/SIG/>
- ❑ Archives des réunions annuelles
précédentes :
<https://members.tei-c.org/Events/meetings/>