

# INTRODUCTION À LA TEI : HISTOIRE

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)

# POURQUOI CODER ?

HISTOIRE

# POURQUOI CODER ?

- Coder = Transformer du langage naturel (langage de tous les jours) en un langage informatique lisible et traduisible par la machine
- Numérique → discipline en pleine émergence, généralisé dans de nombreux domaines
- Aide à la conservation du patrimoine, mise en place de standards informatiques, plus grande stabilité de la donnée
- Favorise le travail collaboratif, à distance, mettant à profit divers domaines de compétence et facilite l'accès de la donnée, sans avoir à se déplacer à l'autre bout du monde
- L'encodage d'une source donne accès à une exploitation plus étendue, plus rapide et plus précise

# INTRODUCTION À LA XML, LANGAGE DE BALISAGE UTILISÉ POUR LA TEI

HISTOIRE

# INTRODUCTION À LA XML, LANGAGE DE BALISAGE UTILISÉ POUR LA TEI

- Qu'est-ce que l'XML ?
- Comment ça fonctionne ?
- XML et ses applications

# QU'EST-CE QUE L'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 imbriqués que l'on désignera comme des parents et enfants
  - Des balises à la suite désignés 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 suivi 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

Exemple correct : `<personne><nom>Martin</nom></personne>`

Exemple incorrect : `<personne><nom>Martin</personne></nom>`

- Attributs : les éléments peuvent être enrichies d'attributs et leurs valeurs

Exemple : 

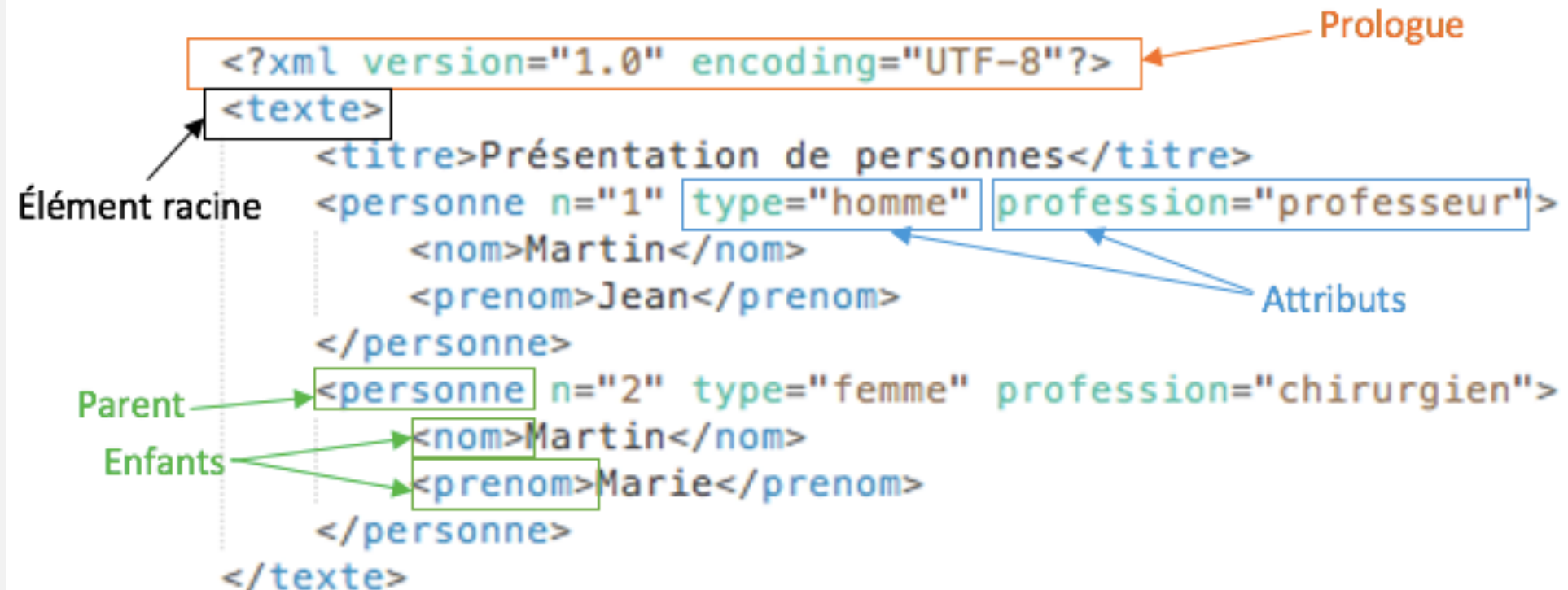
```
<personne n="1" type="homme" profession="professeur">
  <nom>Martin</nom>
  <prenom>Jean</prenom>
</personne>
```

- 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 ?



# XML ET SES APPLICATIONS

- XHTML
  - Langage de balisage hypertexte pour pages web
  - Meilleure syntaxe et encodage plus rigoureux et plus stricte que le HTML
- DocBook
  - Langage de balisage pour la documentation
  - Destiné à l'origine pour des papiers de documentation technique mais généralisé à tout type de livres
- EAD
  - Langage de balisage pour la description archivistique
  - Standard d'encodage des instruments de recherche archivistique, autant en bibliothèques, musées et services d'archive
- TEI
  - Langage de balisage de documents textuels (principalement académiques)
  - Standard d'encodage pour la représentation de textes dans leur format numérique

# INTRODUCTION À LA TEI

HISTOIRE

# INTRODUCTION À LA TEI

- Brève histoire de la mise en place de la TEI
- Les TEI Guidelines
- Consortium collaboratif

# BRÈVE HISTOIRE DE LA MISE EN PLACE DE LA 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
- Janvier 1999-Janvier 2001 : Création du consortium TEI
  - Avoir une organisation officielle qui maintient, développe et promeut 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
  - Version disponible en XML

# 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

## Text Body

- + 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](#)

# 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](#)
- *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
- 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

HISTOIRE

## OUVRAGES

- HAROLD, Eliotte 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 of the Text Encoding Initiative* : <https://journals.openedition.org/jtei/>
- *TEI Special Interest Groups (SIGs)*: <https://tei-c.org/Activities/SIG/>