

La minute du PSIR

19 octobre 2022 - # 56

La TEI (Text Encoding Initiative)



La **TEI** est un standard d'encodage de textes servant à créer des éditions numériques ou imprimées et gérer des corpus de textes pour la recherche en sciences humaines et sociales (SHS). L'acronyme TEI désigne à la fois un cadre conceptuel pour représenter et analyser les textes, un ensemble de règles d'encodage documentées et organisées sous forme de schéma formel (*Guidelines*) et un consortium international pluridisciplinaire.

Un cadre de balisage sémantique et analytique

La TEI fournit un vocabulaire de balises et des règles utilisant le format XML (eXtensible Markup Language). Un document TEI se présente sous la forme d'une arborescence hiérarchique d'éléments délimités par des balises. Il est formé à minima d'un en-tête contenant des métadonnées – pour décrire par exemple les aspects bibliographiques ou matériels du document support – et d'un corps de texte structuré (en sections, chapitres, paragraphes etc., s'il s'agit d'une monographie par exemple). C'est une méthode de balisage "descriptive" par opposition à d'autres langages à balises se focalisant sur le rendu typographique ou la mise en page (comme LaTeX). Un balisage TEI s'attachera par exemple à indiquer qu'une séquence de caractères représente un titre, une expression en langue étrangère, ou une emphase. L'affichage en caractères italiques (ou tout autre mise en forme) est repoussé à une étape ultérieure et peut être défini de manière différenciée pour chaque support de diffusion.

Les Guidelines : une documentation riche, un schéma et des outils

Les solutions d'encodage proposées par la TEI sont organisées au sein de modules spécialisés (éléments communs, théâtre, poésie, sources primaires, dictionnaires, etc.) et décrites sous forme de recommandations génériques. Ces *Guidelines* comprennent ainsi près de 600 balises, avec des déclarations formelles exploitables par les machines ainsi que des définitions en prose à destination des humains. Elles contiennent en outre de très nombreux exemples issus de projets réels. Elles fournissent également des schémas formels et des outils (compilateur, feuilles de style, etc.).

Le consortium TEI : une communauté ouverte et engagée

Le consortium TEI est l'association internationale regroupant des membres institutionnels et individuels. Il prend en charge la maintenance de l'infrastructure technique, la mise à jour des *Guidelines* et fédère sa communauté d'utilisateurs autour d'une conférence annuelle, d'une revue (le journal de la TEI) et d'une liste de discussion ouverte à tous (TEI-I).

Une référence incontournable pour l'édition des sources en SHS

Le standard TEI est devenu une référence incontournable pour l'édition de sources en SHS. Il permet d'obtenir des sources exploitables à un niveau très fin, tout en préservant l'unité de la source sans l'atomiser contrairement à une base de données de type relationnelle. Certaines communautés développent des déclinaisons spécialisées comme EpiDoc pour les documents portant des inscriptions ou la MEI pour la musique.

La richesse de sa documentation, la robustesse de son architecture technique, sa flexibilité ainsi que son ouverture en font aussi l'un des meilleurs formats d'archivage pour le long terme en SHS.

Pour aller plus loin :

<http://www.tei-c.org>

NIVEAU DE
DIFFICULTÉ



Projet
Emmanuelle Morlock

PERSONNES
CONCERNÉES

