



Bibliothèques Virtuelles Humanistes



Encoder l'imprimé

Encodage sémantique

Encoder sémantiquement : pourquoi ?

Par amour de l'encodage ?

Pour mettre en valeur le sens de mon texte ?

Pour générer automatiquement des index :
noms de personne, de lieux, etc... ?

Pour améliorer la finesse des requêtes dans les
outils de fouille de textes ?

Encoder des noms propres

Une balise très générique est disponible en TEI, qui signifie «nom propre» : il s'agit de la balise

`<name>...</name>`

Actuellement, les BVH encodent les noms de personne au moyen de cette balise mais nous allons voir aujourd'hui l'ensemble des balises disponibles dans la version P5 de la TEI afin d'encoder les noms propres.

Encoder les noms de personne

<persName> : nom de personne

<forename> : prénom

<surname> : nom de famille, patronyme

<roleName> : titre officiel, rang, rôle

<nameLink> : particule

<genName> : attribut générationnel (Pline «le Jeune», Henri «IV»)

<addName> : nom additionnel comme un surnom ou un alias

ex : Geoffroy Ier Grisegonelle

```
<persName>  
  <forename>Geoffroy</forename>  
  <genName>Ier</genName>  
  <addName>Grisegonelle</addName>  
</persName>
```

Exercice sur les noms propres

Ouvrir [ex_Semantique.xml](#).

Encoder les noms de personne avec la balise `<persName>` et ses sous-balises :

`<forename>`, `<surname>`, `<nameLink>`,
`<genName>` et `<roleName>`

Ne vous occupez pas des noms de lieux.

Vous avez la possibilité de typer les `persName` avec `@type="mythologie"` ou `@type="historique"` par exemple (comme toujours, vous avez le choix des valeurs qui vous semblent les plus pertinentes)

Encoder le nom d'une organisation

<orgName> sert à encoder le nom d'une organisation ou d'une institution.

Ex : Chapitre de Saint Léon

```
<orgName>  
  Chapitre de  
  <persName>  
    <roleName>Saint</roleName> Léon  
  </persName>  
</orgName>
```

Encoder les noms de lieux

<placeName> : nom de lieu absolu ou relatif, c'est à dire nom d'un bâtiment, village, ville ou pays.

ex : Église Saint-Christophe à Tours

```
<placeName type="eglise">  
  Église  
  <persName>  
    <roleName>Saint</roleName>-Christophe  
  </persName>  
</placeName>  
à  
<placeName type="ville">Tours</placeName>
```

Encoder des noms de lieux (2)

<geogName> : nom ayant une caractéristique géographique (La Loire, le Mont Sinai...)

Cette balise est associée à <geogFeat> qui sert à encoder le nom commun caractérisant le nom géographique (mont, fleuve, vallée, montagne...)

Ex : le Mont Sinai

```
<geogName>  
  Le  
  <geogFeat>Mont</geogFeat>  
  Sinai  
</geogName>
```

Exercice sur les noms de lieux

Reprenez l'exercice [ex_Semantique.xml](#) et

Encodez les noms de lieux avec les balises

- `<placeName>` avec son attribut `@type`
 - `@type="ville"`
 - `@type="pays"`
 - `@type="continent"`
 - `@type="region"....`
- `<geogName>` et sa sous-balise `<geogFeat>`.

Encoder les noms géopolitiques actuels

Nous n'utilisons pas ces balises dans le corps de nos textes anciens mais dans le `teiHeader` :

`<country>` : pays

`<settlement>` : ville

`<region>` : région ou département

Ce sont des balises adaptées à notre découpage administratif actuel.

Dates normalisées

Ces attribut accompagnent la balise <date> (corps du texte) ou <docDate> (page de titre). Leurs valeurs sont normalisées selon la norme iso : aaaa-mm-jj.

@when : Date exacte.

ex : <date when="2012-10-02">Deux octobre 2012</date>

ou <date when="1552">M.D.LII</date>

@notBefore : indique la date la plus antérieure possible

ex : c'était

<date notBefore="2001">un ou deux ans</date> avant sa naissance en <date when="2003">2003</date>.

@notAfter : indique la date la plus postérieure possible

ex : Nous partirons d'ici <date notAfter="2012-03-31">deux mois maximum</date> à partir de ce <date when="2012-01-31">jour</date>.

Ces deux attributs peuvent bien entendu être utilisés ensemble.

L'encodage éditorial

Rappels sur la balise <choice> : permet de proposer une forme alternative du mot encodé.

On utilise avec elle les sous-balises :

<abbr> / <expn> : forme abrégée / désabrégée

<sic> / <corr> : coquille / correction

Lorsqu'on propose une correction, il est en général attendu une mention de responsabilité par le biais de l'attribut @resp (personne) ou @source (document)

ex : <choice><sic>restabit</sic><corr source="errata">restablit</corr></choice>

<orig> / <reg> : forme originale / forme régularisée

On peut parfois utiliser ces balises seules par ex :

- une coquille sans proposition de correction :

les <sic>moustaces</sic>

- une abréviation qu'on ne développe pas :

<abbr>f.</abbr> 5

Exercice Final

Encoder sémantiquement le texte de
l'[ex_Final_Sem.xml](#)

<persName>

<forename>, <surname>, <roleName>,
<addName>, <nameLink>, <genName>

<placeName>

@type

<geogName>

<geogFeat>

<date>

@when, @notBefore, @notAfter

Merci !

Les exercices et supports seront déposés sur Célène dans la semaine.

Des questions sur l'encodage ?