

Numérisation du fonds oulipien: démarches, défis, désirs ...

Lou Burnard

Un mot de notre sponsor



L'importance d'être explicite

L' exploitation informatique nécessite une information

- explicite
- univoque

L' exploitation partagée nécessite une information explicite qui se base sur

- des standards communautaires
- des technologies ouvertes

Un texte n'est pas un document...

Un document est un objet physique, existant dans le monde physique. Un manuscrit, une peinture, une réalisation musicale, une copie spécifique d'un livre...

Un "texte" est quelque chose d'abstrait: la construction d'une communauté de lecteurs.

L'encodage explicite cette abstraction afin de mieux la gérer

Numériser et encoder...

La numérisation nous propose une image, une représentation d'un objet déjà existant

L'encodage nous permet de représenter l'image des idées résultant de ces représentations

Des idées sur quoi?

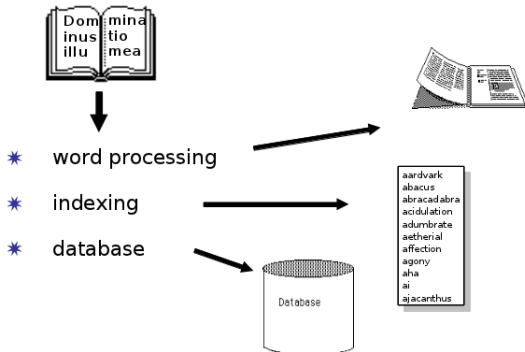
Tout! mais on peut distinguer plusieurs axes...

- infos 'intrinsèques' à l'objet : les formes, couleurs, etc. d'une image; les sons, rythmes, etc. d'une musique; les structures linguistiques (mot, phrase, paragraphe) d'un texte...
- infos 'extrinsèques' ou 'meta' sur l'objet : son type, ses origines, ses buts, ses usages...
- infos 'interprétatives': la portée d'un texte ou d'un dessin, le programme d'une musique ou d'un rite...

distinctions floues mais pervasives...

Pareillement, pour le soft

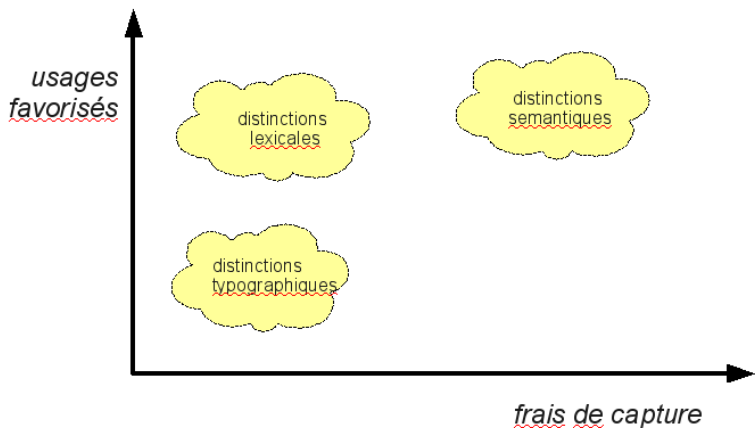
Three kinds of software



(Burnard 1989)

Comment choisir?

On n'encode pas pour se faire plaisir...



Comment structurer l'encodage ?

Il existe (quelques) méthodes classiques d'analyses de données.
L'important c'est de comprendre :

- toute méthode ne serait qu'une modélisation
- il ne faut absolument pas laisser cette tâche aux informaticiens!

Après modélisation, on peut faire une implémentation informatisée; sans modèle, une implémentation risque d'être incompréhensible et aléatoire, inutile..

Analyse des données classique

On identifie...

- les "objets d'intérêt"
- leurs attributs et propriétés
- les relations entre objets
- les procédures/traitements essentiels envisagés

Ce travail est prévu pour la 2ème session de ce séminaire, à Oxford en février 2014

Resultats souhaités

- identification des aspects les plus signifiants des documents du fonds
- propositions d'encodage de ces aspects en se servant du système TEI-XML
- une experimentation sur quelques documents encodés enTEI afin de valider ces propositions

Pourquoi TEI-XML ?



Le buffet de la TEI

TEI vous propose un ensemble de 400+ balises :

- metadonnées (l'entete TEI)
- la structuration basique des textes
- les textes oraux
- la transcription des sources originales
- l'édition critique
- les dictionnaires
- l'annotation linguistique
- les noms de lieux et de personnes et autres entités nommés
- la description de manuscrits
- les liens et analyses hypertextuels
- etc etc
- Et cetera: voir

<http://www.tei-c.org/P5/Guidelines/>



L'esprit TEI

Qu'est-ce que cela veut dire: «être conforme» à la TEI ?

- une pratique consensuelle de balisage
- un lexique commun
- un respect de l'autonomie
- une documentation spécifique au projet et normalisé (ODD)

La standardisation ne devrait pas signifier «fais comme moi»; mais plutôt «explique-moi ce que tu fais. »

Des références...

Site web du Consortium TEI <http://www.tei-c.org>

Dépôt sourceforge <http://tei.sf.net>

Quelques références francophones:

- tei-fr@cru.fr
- <http://meet.tge-adonis.fr>
- <http://lespetitescases.net/index102/>
- <http://www.culture.gouv.fr/culture/dglf/riofi/tei.htm>
- [http://artist.inist.fr/article.php3?id_article=122"/](http://artist.inist.fr/article.php3?id_article=122/)