

Python pour le traitement des sources XML

Matthias GILLE LEVENSON

AMU – CIHAM UMR 5648

prénom [point] gille [point] levenson [at] ens-lyon.fr

2 octobre 2023



1 Plan de la séance

- Une partie de présentation technique, méthodologique et philologique
- Une partie de présentation de l'interface
- Une partie de travaux pratiques avec entraînement d'un modèle

lxml et la non-récurtivité

- Le traitement des données XML via python ne suit pas la logique d'imbrication du XML
- Il y aura donc des cas où python ne sera pas adapté (en général, tout ce qui a trait à la production d'éditions et à la transformation complexe de sources XML)
- Python sera bien adapté au traitement complexe de documents XML qui nécessitent l'utilisation de listes, dictionnaires
- De même, l'intérêt principal de Python est la possibilité d'utiliser des librairies et outils externes de traitement (traitement de l'image, TAL).

Limites de lxml

- Limité à XPath 1.0, compensé par la puissance de python

Les espaces de nom



Parser l'arbre XML

- La méthode 'parse()'
- Dans le cas d'inclusions, on doit explicitement demander à ce que les sources externes soient aussi parsées

Naviguer dans l'arbre



Récupérer les noeuds textuels



Créer un élément



Créer un attribut



Références I
