

# Documentation

- Pour les données au format csv, on utilisera:
  - `pandas.read_csv`. Voir [dataframes pandas](#). Pandas permet également de lire les données au format xls et xlsx (Excel).
  - Pandas permet de manipuler des données de type DataFrame. Si d est votre DataFrame et que vous avez besoin de passer au format array de numpy : `m = d.as_matrix()`
- Pour les données au format xml, on utilisera la librairie etree. L'énoncé suivant donne les indications pour utiliser cette librairie [Lecture des fichiers xml](#)
- Pour le format json, voir : [Lecture des fichiers json](#)

## Datasets

### Bases de textes

#### Anglais

- [web-Amazon.html](#)
- [bbc.html](#)

#### Français

- [Articles de l'Est Republicain](#)
- [Base Frantext](#)
- [Flux d'actualités officielles](#)
- [Associations loi 1901 \(2014\)](#)
- [Extraire un flux twitter en francais](#)

### Recommandation

- [|Recommandation musicale](#)
- [Recommandation littéraire](#)
- [Recommandation de blagues](#)
- [|Recommandation culinaire \(!!\)](#)

### Analyse

- [fr-esr-enseignants-titulaires-esr-public-national@mesr](#)
- [parrainages-valides-elections-presidentielle-2017@public](#)

- [fete-de-la-musique-2016](#)

## Divers

- [Prénoms français \(2002 à 2012\)](#)

## Ressources

- [Stanford Information Retrieval](#)
- [index.html](#)
- [Centre National de Ressources Textuelles et Lexicales](#)
- [SNAP](#)

From:

<https://wiki.centrale-med.fr/informatique/> - **WiKi informatique**

Permanent link:

[https://wiki.centrale-med.fr/informatique/public:omi-5a-o-rech:4.\\_travail\\_final](https://wiki.centrale-med.fr/informatique/public:omi-5a-o-rech:4._travail_final)

Last update: **2018/02/22 10:22**

