

Nettoyer et transformer des données avec OpenRefine : des premiers pas aux usages avancés

By: bouchard , Comments: 0

[Ressource] ^[1]

[OpenRefine](#) ^[2] est un logiciel libre et gratuit destiné à nettoyer, préparer, transformer et enrichir des données. Il est utilisable dans de nombreuses contextes scientifiques et dans les métiers de l'information et de la documentation. Son interface est proche d'un tableur et permet de réaliser des opérations au moyen de menus ou en écrivant des formules dans un langage spécifique. Au cours de la formation, les fonctionnalités simples et avancées d'OpenRefine seront présentées de manière progressive et dynamique, en s'appuyant sur des exemples et des exercices pratiques. Des supports et éléments de documentation en français seront fournis.

Programme de la formation animée par Mathieu Saby (Bibliothèque universitaire de l'Université de Nice Sophia Antipolis)

- Présentation du logiciel
- Installation d'extensions
- Import et export de données et de traitements
- Repérage et harmonisation de variantes ou d'anomalies (tris, facettes, clusterisation)
- Modification de l'organisation générale des données
- Croisement de plusieurs fichiers
- Extraction de motifs (introduction aux expressions régulières)
- Scission de cellules contenant plusieurs valeurs, et regroupement de plusieurs valeurs dans une même cellule
- Enrichissement par récupération automatique de données sur le web (API et web scraping)
- Alignement de données avec des référentiels externes (Wikidata...)
- Formules en langage GREL (spécifique d'OpenRefine)
- Formules en langage Python

Lien: [Dossier général sur OpenRefine par Mathieu Saby](#) ^[3]

[Supports de la formation du 25/05/2018 et données](#) ^[4]

[Mémo sur le modèle de données d'Openrefine, GREL et Jython](#) ^[5]

Sujets principaux:

[Traiter l'information](#) ^[6]

Sujet(s) secondaire(s):

[Production et traitement du document numérique](#) ^[7], [Techniques de gestion documentaire](#) ^[8]

Outil(s):

[OpenRefine](#) ^[9]

Types de public:

[Doctorant](#) ^[10], [Enseignant du supérieur, chercheur](#) ^[11], [Professionnels de l'information](#) ^[12]

Type de ressource:

[supports de cours](#) ^[13]

Date de création:

21/06/2018

Date de mise à jour:

21/06/2018

URL source: <http://urfist.chartes.psl.eu/ressources/nettoyer-et-transformer-des-donnees-avec-openrefine-des-premiers-pas-aux-usages-avances>

Liens

[1] <http://urfist.chartes.psl.eu/actualites/ressource>

[2] <http://openrefine.org/>

[3] <https://osf.io/uf98g/>

- [4] <https://osf.io/uqwrp/>
- [5] <https://osf.io/aew57/>
- [6] <http://urfist.chartes.psl.eu/vocthematiques0/traiter-linformation>
- [7] <http://urfist.chartes.psl.eu/vocsujetssecondaires0/production-et-traitement-du-document-num%C3%A9rique>
- [8] <http://urfist.chartes.psl.eu/vocsujetssecondaires0/techniques-de-gestion-documentaire>
- [9] <http://urfist.chartes.psl.eu/outils/openrefine>
- [10] <http://urfist.chartes.psl.eu/types-de-public/doctorant>
- [11] <http://urfist.chartes.psl.eu/types-de-public/enseignant-du-sup%C3%A9rieur-chercheur>
- [12] <http://urfist.chartes.psl.eu/types-de-public/professionnels-de-l%E2%80%99information>
- [13] <http://urfist.chartes.psl.eu/ressources/supports-de-cours>