

Journées Istex – 16 et 17 juin 2025

Atelier 1.1.3
Formation de formateurs·rices
dans les bibliothèques

Sabine Barreaux et Emeline Caule – Inist
Claire Tignolet – Urfist Paris

Objectif de l'atelier

Cet atelier permettra de réfléchir à la manière de mobiliser des relais dans les établissements pour mieux faire connaître les services de l'infrastructure Istex dans les différentes communautés scientifiques et encourager leur utilisation auprès des chercheurs et des personnels d'appui à la recherche.

Il s'agira plus particulièrement d'échanger avec vous sur des questions relatives à l'identification des personnes pouvant être concernées par ce rôle de relais et sur les moyens à mettre en œuvre pour former ces personnes et les accompagner au mieux dans cette démarche.

Plan de l'atelier

- Présentation d'Istex
- Tour de table
- Formations Istex par l'Inist
- Échanges

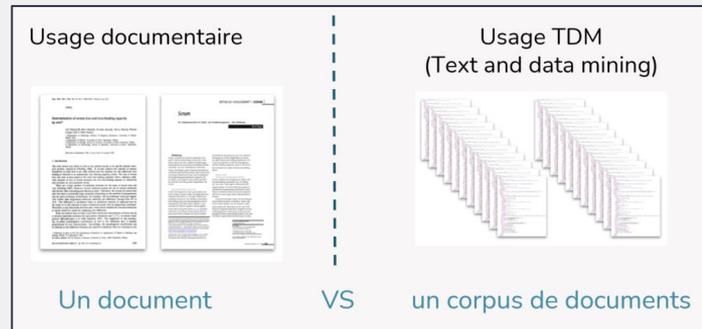
Présentation d'Istex

Istex : contenu

- Initiative d'excellence en Information Scientifique et Technique : “Construire le socle de la bibliothèque scientifique numérique nationale”
- Acquisition massive et centralisée d'archives scientifiques
 - Issue des Licences Nationales
 - Collections rétrospectives multilingues et multi disciplinaires
- Mise à disposition des données : Plateforme nationale (Inist-CNRS)
- 30 064 284 documents
 - 50 bouquets éditeurs négociés
 - 10 212 revues
 - de 1455 à 2023

Istex : 2 types d'usage

- Usage documentaire
 - Consultation, lecture, citation d'un document
 - Accès : Google Scholar, Click&Read, solution de découverte, Revue de sommaire, Istex Search
- Usage TDM (fouille de textes)
 - Exploitation et traitement d'un lot de documents par des outils spécialisés
 - Accès : Corpus scientifiques, Lodex, Loterre, Istex TDM, Istex Search



Les services Istex

Interrogation et téléchargement

Istex Search
Constitution de corpus

Visualisation

Lodex
Exploration du corpus

Exploration et enrichissement

Services TDM
Fouille de textes (web services)

Lodex
Analyse de corpus

Loterre
Terminologies

Corpus scientifiques – Cas d’usages

Des exemples de corpus avec des cas d’application ou d’utilisation d’outils de TDM
> <https://www.istex.fr/constitution-de-corpus/exemples-de-corpus/>

Des collections de corpus offrant des exemples d’exploitation des ressources Istex dans des contextes scientifiques variés
> <https://corpus.istex.fr/>

Tour de table

Présentation des participants

- Qui êtes-vous ? (établissement, activités principales)
- Quelle est votre connaissance et votre pratique des services Istex ?
- Dans quel cadre envisagez-vous de former à Istex ?

Les formations Istex par l'Inist

Les formations Istex (1/3)

- Panorama des formations
 - 48 formations depuis 2017
 - Env. 630 participants
 - Présentiel et/ou distanciel
 - Plusieurs profils : professionnel.les IST, doctorant-es, chercheur-euses, ingénieur-es
 - Contextes : Urfist, ENSSIB, ANF, Écoles doctorales, universités, Media (CRFCB)
- En parallèle des formations, des **présentations** à Istex de 1h à 3h sont réalisées selon les besoins exprimés : présentation globale, LODEX, TDM...
- Des **webinaires** d'1h env. sans inscription sont également programmés plusieurs fois dans l'année (dernier thème proposé : les nouvelles fonctionnalités d'Istex Search)

Les formations Istex (2/3)

Formules des formations + objectifs

Formules spécifiques à Istex

- 1 journée complète (**Media**)
 - présentation Istex et constitution de corpus
 - atelier pratique pour manipuler les outils présentés (Istex Search et Lodex)
 - objectif : initier aux pratiques de constitution de corpus
- 2 ½ journées espacées (**écoles doctorales**)
 - présentation Istex et constitution de corpus
 - atelier pratique pour manipuler les outils présentés (Istex Search et Lodex)
 - objectif : montrer l'intérêt d'Istex comme ressource documentaire et matériau pour la recherche

Formules Fouille de textes

- 1 journée complète (**Urfist, Enssib-DCB**)
 - présentation Istex, constitution de corpus et introduction au TDM
 - ateliers pratiques sur outils TDM (Gargantext, web services TDM)
 - objectif : découverte d'Istex pour le TDM
- événements dédiés à la fouille de textes sur plusieurs jours (**Enssib, ANF-TDM**)
 - 1 plénière d'1h à 2h sur Istex pour présenter la constitution de corpus avec des démos
 - des ateliers pratiques sur des outils de TDM
 - objectif : découverte du TDM

Les formations Istex (3/3)

Exemple d'une formation d'1 journée

- Durée : 6h
- 1 à 3 formateur·rices
- Contenu :
 - Présentation d'Istex
 - Fouille de textes / TDM (*optionnel*)
 - Constitution et analyse d'un corpus documentaire
 - TP (interrogation, téléchargement, visualisation, enrichissements)



TP - SOCIOLOGIE DE L'ALIMENTATION

Formation à destination des URFIST

1 Description générale

Besoin "Je suis sociologue et je débute une recherche sur l'évolution des motivations des personnes véganes et végétariennes au fil du temps (sociologie de l'alimentation)."

Objectif Détection des études sur le véganisme et le végétarisme. Analyse avec Lodex.

Outils de TDM envisagés par la suite Lodex et les web services de TDM.

Contraintes imposées par les outils La plupart des web services ne traitent qu'une langue à la fois (conserver la langue majoritaire).

2 Étapes

2.1 Interrogation de la base Istex

Étape 1 Se rendre sur Istex Search <https://search.istex.fr/>.

Étape 2 Rechercher les formes végane, végétarien, végétarisme, vegan, vegetarian et vegetarianism.

Étapes 3 Pour limiter le silence, rechercher les dérivés et variantes morphosyntaxiques (féminin et pluriel).

Étape 4 Pour limiter le bruit, on peut préciser les champs les plus à même de renvoyer des résultats pertinents. La liste des champs est accessible dans la recherche assistée. L'objectif est de rechercher les mots-clés dans le **titre**, le **résumé** et les **mots-clés d'auteurs**.

Étape 5 En affinant ma problématique, je découvre que le terme **vegan** a été créé en 1944 par le cofondateur de la Vegan Society, Donald Watson. Je souhaite vérifier si cette revendication a eu un impact social et s'il transparaît dans la littérature scientifique. Peut-on limiter les dates de publication du corpus à des documents publiés après 1944 ? Quel champ interroger ?

Étape 6 Je souhaite disposer d'un résumé pour chacun des documents pour m'assurer de la pertinence des articles avant une lecture complète. Comment m'assurer de la présence de résumés dans les documents ?

Étape 7 Quelle est la langue majoritaire ? Conditionner la requête pour obtenir un corpus monolingue.

2.2 Téléchargement du corpus

Étape 8 Télécharger les documents au format approprié.



2.3 Chargement dans Lodex

Étape 9 Se rendre sur votre instance Lodex, se connecter et charger les données. Le lien, votre identifiant et votre mot de passe sont disponibles depuis le site contenant les supports <https://urfist-paris-1591.formation.lodex.fr/instance/suports-formation>.

- Aller dans l'interface administrateur en cliquant sur **Voir plus > Admin** en bas à droite.
- Importer le corpus en glissant le fichier .zip téléchargé **sans décompression préalable**.
- Choisir un loader¹.
- Cliquer sur **Importer les données**. Comment vérifier que votre importation est complète ?

Étape 10 Créer votre site.

- Importer le modèle² fourni (fichier *URFIST-modele-vegan.tar*) : menu en haut à droite : > **Modèle** > + **Importer un modèle**.
- Publier le site pour une première visualisation en cliquant sur **Publier** en haut à droite.
- Cliquer sur l'icône en forme d'œil pour voir le résultat.
- Explorer les différents graphiques à partir de l'onglet **Graphiques** en bas.
- Consulter quelques ressources grâce à l'onglet **Recherche**.
- Modifier un champ du modèle : la date de création. Aller sur l'interface administrateur en bas à droite **Voir plus > Admin > Affichage (onglet en haut à droite de Données) > Page d'accueil**.
- Cliquer sur le champ à modifier (sur la roue crantée) et aller modifier le texte dans **Valeur arbitraire** puis sauvegarder.

2.4 Premiers pas en TDM

Étape 11 Extraire les mots-clés des résumés via le web service *Teefi*.

- Aller dans **Données > Enrichissements** et cliquer sur + **Ajouter**.
- Donner le nom **Mots-clés (WS)**, aller chercher l'URL du web service *Teefi* dans le catalogue en cliquant sur le bouton vert à droite du champ **URL du web service**. Selon votre corpus, quelle version de *Teefi* faut-il sélectionner ? Choisissez le nombre de termes à extraire.
- Choisir **Résumé** dans la **Colonne de la source**, cliquer sur **Sauvegarder**. Cliquer sur **Lancer**. Observer la sortie du web service dans le tableau des données. Que constatez-vous ?

Dans Lodex, avant de faire un graphique, il est nécessaire de déclarer la colonne comme une ressource (c.-à-d. créer un champ qui contient les données utilisées dans le graphique, dans ce cas : le résultat de Teefi).

¹ Un loader est un script d'adaptation du fichier à Lodex. Il dépend du format de fichier fourni en entrée.

² Pour rappel, un modèle est un fichier .tar qui permet de mettre en forme le site créé avec Lodex.



- d) Aller dans *Affichage > Ressource principale*, cliquer sur *+ Nouveau champ*. Pour paramétrer ce nouveau champ : dans *Étiquette* nommer le champ *Mots-clés*, sélectionner *Colonne(s) existante(s)* et aller chercher la colonne nommée *Mots-clés (WS)* (le contenu de la colonne s'affiche à droite).
- e) Après avoir sélectionné la colonne, cliquer sur *Ajouter une opération* pour appliquer l'opération de transformation *GET*, renseigner *term* dans le *path*. À qui sert cette dernière manipulation ? Dans *Affichage* choisir le format *Texte-Liste de valeurs*. *Confirmer* puis *Sauvegarder*.
- f) Pour créer le graphique, aller dans *Affichage > Graphiques* cliquer sur *+ Nouveau champ*, nommer le graphique en renseignant *Mots-clés les plus spécifiques* dans le champ *Étiquette*. Choisir la routine³ *distinct-by* puis choisir le champ sur lequel la routine va s'appliquer grâce au menu déroulant (soit le champ *Mots-clés*). Dans *Affichage*, choisir le format *Graphique à bulles* en filtrant les résultats : dans *Paramètres des Données* mettre valeur minimum à afficher à 2 et choisir le tri *Descendant*. Cliquer sur *Confirmer* puis *Sauvegarder*.
- g) Pour créer un filtre avec ces nouvelles valeurs, se rendre dans *Affichage > Recherche* et facettes et cocher la case *Mots-clés* dans les *Facettes*.
- h) Enfin, aller consulter les résultats de votre premier traitement de TDM . *Combien de documents contiennent le mot-clé meat consumption ?*

2.5 Pour aller plus loin...

Étape 12 Extraire les entités nommées des résumés pour déterminer les noms de lieux les plus fréquents.

- a) Utiliser le web service *entityTag - Extraction de différentes entités nommées d'un texte en anglais* sur les résumés, le nommer *entityTag* (au besoin s'aider des détails de l'étape 11). Observer la sortie du web service dans le tableau des données. Que constatez-vous ?
- b) Pour créer une colonne contenant uniquement les noms de lieux, aller dans *Données > Enrichissements* et cliquer sur *+ Ajouter*. Donner le nom *ENgeographie*, cliquer sur *Mode avancé*⁴ et coller le code ci-après, cliquer sur *Sauvegarder*. Cliquer enfin sur *Lancer*. Pour des raisons de temps de calcul (et temps de TP) ce code ne conserve que le premier nom de lieu du document grâce à l'instruction `.slice(0,1)`.

3. Les routines de Lodex sont des scripts, fournis par Lodex ou stockables n'importe où sur le web, qui peuvent effectuer des agrégations, des calculs, des reformatages et renvoyer des données sous une forme utilisable (souvent un format de type graphique).

4. La mode avancé offre une meilleure flexibilité pour transformer les données. Il s'agit de code en Lodash, une librairie Javascript (cf. tutoriel). Il n'est pas toujours nécessaire de comprendre le code pour appliquer des modifications : les transformations les plus usuelles sont disponibles via ce lien.



```
{assign}
path = value
value = get("value.entityTag.LDC").slice(0,1)
```

- c) Faire tourner le web service *Associer un terme au vocabulaire Pays et Subdivisions* sur la nouvelle colonne *ENgeographie* en créant de nouveau un enrichissement nommé *Lieux*. Ce web service permet d'obtenir un code qui permet de créer une cartographie.
- d) Aller dans *Affichage > Ressource principale*, cliquer sur *+ Nouveau champ*. Pour paramétrer ce nouveau champ : dans *Étiquette* nommer le champ, sélectionner *Colonne(s) existante(s)* et aller chercher la colonne nommée *Lieux*. Après avoir sélectionné la colonne, cliquer sur *Ajouter une opération* pour appliquer une transformation *GET*, renseigner *cartographyCode* dans le *path*. Cliquer sur *Sauvegarder*.
- e) Pour créer le graphique : aller dans *Affichage > graphiques* cliquer sur *+ Nouveau champ*, nommer le graphique en renseignant *Lieux cités* dans le champ *Étiquette*. Choisir la routine *distinct-by* puis choisir le champ sur lequel la routine va s'appliquer grâce au menu déroulant. Enfin, dans *Affichage*, choisir le format *Graphique* puis *Cartographie* et paramétrer la carte (nombre maximum de champs : 10000; choisir une couleur; afficher l'info-bulle) et sauvegarder. *Quels sont les pays les plus cités ? Pourquoi ?*

Ce site contient la correction du TP : <https://urfist-paris-1591.formation.lodex.fr/instance/correction-tp-vegan>.

Échanges sur les formations Istex en établissement

1. Quelles formations par les établissements ?

Faire le point sur votre contexte, vos besoins et vos attentes en tant que formateur·rice à Istex

- Quels seraient les objectifs principaux des formations que vous envisagez ?
- A quel(s) besoin(s) prioritaire(s) ou usage(s) pourraient-elles répondre ?
- Décrire la ou les formations que vous envisagez :
 - Modalités / Public / Durée / Cadre / Objectif ou contenu clé

2. Par quels formateurs ou formatrices ?

Identifier et valoriser les futur·es formateur·rices

- Qui serait en charge de formations Istex dans vos structures ?
- Comment valoriser ou reconnaître leur activité ?
- Quels liens auraient-ils·elles...
 - avec l'équipe Istex de l'Inist ?
 - avec les contacts Istex en établissement ?
 - avec les chercheur·euses de l'établissement ?

3. Comment former ces formateur.trice.s ?

Proposer des supports et méthodes d'accompagnement aux formateur·rices

- Au-delà de l'existant, quels dispositifs seraient à privilégier ?
- Quel matériel pédagogique faudrait-il développer en priorité ?
- Avec quels acteurs ?

Les sources d'informations

- Se tenir informé·e
 - Article d'actualité : <https://www.istex.fr/category/actualites/>
 - Revues de l'équipe
- Chercher de l'aide / Contribuer à l'amélioration d'Istex
 - Contact
 - Via le formulaire : <https://www.istex.fr/contact/>
 - Via la liste : contact@listes.istex.fr
 - Liste de discussion Istex : users@listes.istex.fr
 - Liste de discussion Lodex : <https://groupes.renater.fr/sympa/info/lodex>

Merci pour votre participation