



# Corpus Réfugié·es

“Je suis doctorant·e en sociologie, je débute une thèse sur les réfugié·es à travers le monde au 21e siècle. Je souhaite construire une bibliographie sur le sujet, distinguer la part de documents traitant des réfugié·es climatiques et des réfugié·es politiques et enfin identifier leur origine géographique.”

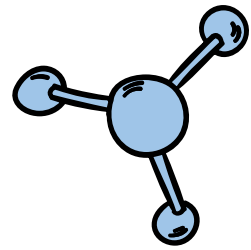
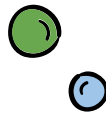
# Description générale

Pour répondre à cette problématique, il faut constituer un corpus de documents traitant des réfugié·es grâce au réservoir Istex. Ce corpus va ensuite être exploré et enrichi dans Lodex.

**Objectif** : repérer les études sur les réfugié·es et débiter leur analyse.

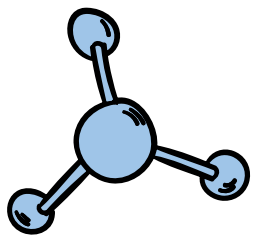
**Outil de TDM utilisé** : Lodex et les web services associés.

**Contraintes imposées par l'outil** : Istex propose un format de sortie adaptée à Lodex. Pour utiliser les web services qui m'intéressent, il faut vérifier que les documents sont en anglais et contiennent des résumés.



## Exercice 1

Construire une requête  
Istex



# Étape 1

Se rendre sur [Istex Search](#).

# Étape 1

Se rendre sur [Istex Search](#).

# Étape 2

Rechercher les formes anglaises et françaises *refugee*, *réfugié*, *asylum seeker* et *demandeur d'asile*.



◦◦. Combien de ressources sont retournées par Istex ?

# ISTEX Search

Créez et téléchargez votre corpus scientifique  
Première étape de votre projet de fouille de textes

## Résultats de votre requête



76 123 documents trouvés

refugee réfugié "asylum seeker" "demandeur d'asile"

RECHERCHER

Requête brute : [https://api.istex.fr/document?q=refugee réfugié "asylum seeker" "demandeur d'asile"&size=10&from=0&rankBy=qualityOverRelevance&ou...](https://api.istex.fr/document?q=refugee+r%C3%A9fugi%C3%A9+%22asylum+seeker%22+%22demandeur+d%27asile%22&size=10&from=0&rankBy=qualityOverRelevance&ou...)

### BOUQUET (48)

Rechercher

- wiley 18 698
- oup 12 631
- elsevier 8 973
- cambridge 8 334

### Indicateurs sur votre corpus (76 123 documents)

Résumé



32140 doc.

PDF texte



73514 doc.

Texte nettoyé



34661 doc.

Langue de publication



- anglais : 73043 doc. (96 %)
- français : 2724 doc. (3,6 %)
- allemand : 279 doc. (0,4 %)
- autres : 74 doc. (0,1 %)

## Étape 3 : Limiter le bruit et le silence

## Étape 3.1 : supprimer le silence

- Rechercher les variantes (singulier / pluriel et féminin / masculin) de chacun des termes de la requête.



- ○ ○ Cette requête a permis de rajouter des documents. Combien ?  
Que risque-t'on si l'on constitue un corpus sans cette étape ?



# ISTEX Search

Créez et téléchargez votre corpus scientifique  
Première étape de votre projet de fouille de textes

## Résultats de votre requête

141 352 documents trouvés



refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile"  
"demandeuse d'asile" "demandeuses d'asile"

RECHERCHER

Requête brute : [https://api.istex.fr/document?q=\(refugee+refugees+réfugié+réfugiée+référugés+référugées+\"asylum+seeker\"+\"asylum+seekers\"+\"demandeur+d'asile\"+\"demandeurs+d'asile\"+\"demandeuse+d'asile\"+\"demandeuses+d'asile\"\)](https://api.istex.fr/document?q=(refugee+refugees+réfugié+réfugiée+référugés+référugées+\)

### BOUQUET (47)

Rechercher

- wiley 32 071
- oup 20 188
- cambridge 19 133
- elsevier 16 599

### Indicateurs sur votre corpus (141 352 documents)

Résumé



60605 doc.

PDF texte



136572 doc.

Texte nettoyé



63272 doc.

Langue de publication



- anglais : 132313 doc. (93,6 %)
- français : 8075 doc. (5,7 %)
- allemand : 780 doc. (0,6 %)
- autres : 180 doc. (0,1 %)

# ISTEX Search

Créez et téléchargez votre corpus scientifique  
Première étape de votre projet de fouille de textes

## Résultats de votre requête



142 595 documents trouvés

refugee\* réfugié\* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile"

RECHERCHER

Requête brute : [https://api.istex.fr/document?q=refugee\\* réfugié\\* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile"](https://api.istex.fr/document?q=refugee*%20r%C3%A9fugi%C3%A9*%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d%27asile%22%20%22demandeurs%20d%27asile%22%20%22demandeuse%20d%27asile%22%20%22demandeuses%20d%27asile%22)



### BOUQUET (49)

Rechercher

<input type="checkbox"/> wiley	32 125
<input type="checkbox"/> oup	20 241
<input type="checkbox"/> cambridge	19 157
<input type="checkbox"/> elsevier	16 641

### Indicateurs sur votre corpus (142 595 documents)

Résumé



61226 doc.

PDF texte



137791 doc.

Texte nettoyé



64052 doc.

Langue de publication



● anglais : 133540 doc. (93,6 %)  
● français : 8088 doc. (5,7 %)  
● allemand : 783 doc. (0,5 %)  
● autres : 180 doc. (0,1 %)

## Étape 3.2 : Limiter le bruit

- Pour supprimer le bruit, on peut préciser les champs les plus à même de renvoyer des résultats pertinents (soit le *titre*, le *résumé* et les *mots-clés d'auteur·ices* dont la dénomination est *title*, *abstract* et *subject.value*). La requête prend alors la forme *champ:()*. La liste des champs est accessible dans la recherche assistée.
- Pour limiter le bruit, on doit s'assurer de ne sélectionner que des articles de recherche (filtre : *Type de contenu* ; nom technique : *genre*).



○ ○ ○ Après suppression du bruit, combien de ressources écrites en français sont retournées ?

# ISTEX Search

Créez et téléchargez votre corpus scientifique  
Première étape de votre projet de fouille de textes

## Résultats de votre requête



3 667 documents trouvés

(title:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") abstract:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") subject.value:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile")) AND genre.raw:"research-article"

RECHERCHER

Requête brute : [https://api.istex.fr/document?q=\(title:\(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur...](https://api.istex.fr/document?q=(title:(refugee%20refugees%20r%C3%A9fugi%C3%A9%20r%C3%A9fugi%C3%A9e%20r%C3%A9fugi%C3%A9s%20r%C3%A9fugi%C3%A9es%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d%27asile%22%20%22demandeurs%20d%27asile%22%20%22demandeuse%20d%27asile%22%20%22demandeuses%20d%27asile%22)%20abstract:(refugee%20refugees%20r%C3%A9fugi%C3%A9%20r%C3%A9fugi%C3%A9e%20r%C3%A9fugi%C3%A9s%20r%C3%A9fugi%C3%A9es%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d%27asile%22%20%22demandeurs%20d%27asile%22%20%22demandeuse%20d%27asile%22%20%22demandeuses%20d%27asile%22)%20subject.value:(refugee%20refugees%20r%C3%A9fugi%C3%A9%20r%C3%A9fugi%C3%A9e%20r%C3%A9fugi%C3%A9s%20r%C3%A9fugi%C3%A9es%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d%27asile%22%20%22demandeurs%20d%27asile%22%20%22demandeuse%20d%27asile%22%20%22demandeuses%20d%27asile%22))%20AND%20genre.raw:%22research-article%22)

### BOUQUET (16)

Rechercher

- oup 1 035
- elsevier 650
- sage 597
- cambridge 445

### Indicateurs sur votre corpus (3 667 documents)



# Étape 4

Répondre aux contraintes **scientifiques** et **techniques**



En vous aidant du filtre *date de publication* et du sujet de thèse du cas d'usage, donnez une **contrainte scientifique** non respectée.

## Rappel de la problématique :

“Je suis doctorante en sociologie, je débute une thèse sur les réfugié-es à travers le monde au 21e siècle. Je souhaite construire une bibliographie sur le sujet, distinguer la part de documents traitant des réfugié-es climatiques et des réfugié-es politiques et enfin identifier leur origine géographique.”

# ISTEX Search

Créez et téléchargez votre corpus scientifique  
Première étape de votre projet de fouille de textes

## Résultats de votre requête



2 273 documents trouvés

(title:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") abstract:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") subject.value:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile")) AND genre.raw:"research-article" AND publicationDate:[2000 TO \*]

RECHERCHER

Requête brute : [https://api.istex.fr/document?q=\(title:\(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur...](https://api.istex.fr/document?q=(title:(refugee refugees réfugié réfugiée réfugiés réfugiées \)



### BOUQUET (15)

Rechercher

- oup 590
- elsevier 408
- sage 363
- springer-journals 332

### Indicateurs sur votre corpus (2 273 documents)



- anglais : 2241 doc. (98,6 %)
- français : 24 doc. (1,1 %)
- allemand : 6 doc. (0,3 %)
- autres : 1 doc. (0 %)

# Étape 4

Répondre aux contraintes **scientifiques** et **techniques**

Pour répondre aux **contraintes imposées par l'outil**, il faut s'assurer de la présence de résumés dans les documents et s'assurer qu'ils sont en anglais.



Enfin, quelle est la taille du corpus sélectionné ?



# ISTEX Search

Créez et téléchargez votre corpus scientifique  
Première étape de votre projet de fouille de textes

## Résultats de votre requête

2 036 documents trouvés



RECHERCHER

(title:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") abstract:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") subject.value:(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile")) AND genre.raw:"research-article" AND publicationDate:[2000 TO \*] AND abstract:\* AND language:eng

Requête brute : [https://api.istex.fr/document?q=\(title:\(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile"\) abstract:\(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile"\) subject.value:\(refugee refugees réfugié réfugiée réfugiés réfugiées "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile"\)\) AND genre.raw:"research-article" AND publicationDate:\[2000 TO \\*\] AND abstract:\\* AND language:eng](https://api.istex.fr/document?q=(title:(refugee%20refugees%20réfugié%20réfugiée%20réfugiés%20réfugiées%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d'asile%22%20%22demandeurs%20d'asile%22%20%22demandeuse%20d'asile%22%20%22demandeuses%20d'asile%22)%20abstract:(refugee%20refugees%20réfugié%20réfugiée%20réfugiés%20réfugiées%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d'asile%22%20%22demandeurs%20d'asile%22%20%22demandeuse%20d'asile%22%20%22demandeuses%20d'asile%22)%20subject.value:(refugee%20refugees%20réfugié%20réfugiée%20réfugiés%20réfugiées%20%22asylum%20seeker%22%20%22asylum%20seekers%22%20%22demandeur%20d'asile%22%20%22demandeurs%20d'asile%22%20%22demandeuse%20d'asile%22%20%22demandeuses%20d'asile%22))%20AND%20genre.raw:%22research-article%22%20AND%20publicationDate:%22[2000%20TO%20*]%22%20AND%20abstract:%22*%22%20AND%20language:eng)

### BOUQUET (13)

Rechercher

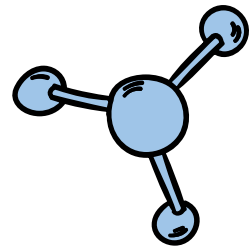
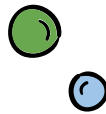
- oup
- elsevier
- ...

493  
384  
240

### Indicateurs sur votre corpus (2 036 documents)

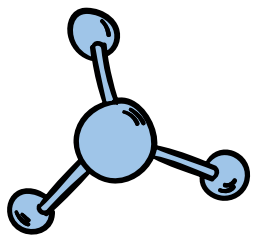






## Exercice 2

### Premiers pas vers le TDM



# Étape 1

Télécharger le corpus : extraire le corpus Réfugié-es en utilisant l'équation corrigée et en choisissant le format adapté pour un import dans Lodex.

## Configurez votre téléchargement



USAGE PERSONNALISÉ

**LODEX**

CORTEXT

GARGANTEXT

NOOJ

- |   |   |  |
|---|---|--|
| <input type="checkbox"/> Texte intégral | <input checked="" type="checkbox"/> Métadonnées | <input type="checkbox"/> Enrichissements |
| <input type="checkbox"/> PDF            | <input checked="" type="checkbox"/> JSON        | <input type="checkbox"/> multicat        |
| <input type="checkbox"/> TEI            | <input type="checkbox"/> XML                    | <input type="checkbox"/> nb              |
| <input type="checkbox"/> TXT            | <input type="checkbox"/> MODS                   | <input type="checkbox"/> grobidFulltext  |
| <input type="checkbox"/> CLEANED        |   | <input type="checkbox"/> refBibs         |
| <input type="checkbox"/> ZIP            | <input type="checkbox"/> Annexes                | <input type="checkbox"/> teeft           |
| <input type="checkbox"/> TIFF           | <input type="checkbox"/> Couvertures            | <input type="checkbox"/> unitex          |

Trier par : pertinence & qualité ▾

Télécharger 2 036 / 2 036

TOUT

Format de l'archive : ZIP ▾

Compression : moyenne ▾

TÉLÉCHARGER

### Lodex

Application web open-source dédiée aux données structurées qui permet de visualiser et d'enrichir ses données puis de les transformer en site web.

[EN SAVOIR PLUS](#)

### Requête

```
(title:(refugee refugees réfugié réfugiée
réfugiés réfugiées "asylum seeker" "asylum
seekers" "demandeur d'asile" "demandeurs
d'asile" "demandeuse d'asile" "demandeuses
d'asile") abstract:(refugee refugees réfugié
réfugiée réfugiés réfugiées "asylum seeker"...
```

### Requête brute complète

```
https://api.istex.fr/document?q=(title:(refugee
refugees réfugié réfugiée réfugiés réfugiées
"asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"...
```

## Étape 2

Importer le corpus dans Lodex.

Cette étape est la plus sensible aux erreurs de manipulations des données. Nous allons donc la faire tous ensemble, nous récapitulons la procédure dans la suite.

## Étape 2

Importer le corpus dans Lodex.

- Se rendre sur votre instance Lodex : se connecter avec votre nom d'utilisateur et votre mot de passe.
- Aller dans l'interface administrateur en cliquant sur *Voir plus > Admin*.
- Importer le corpus en glissant le fichier .zip téléchargé **sans décompression préalable**.
- Choisir le loader *ZIP résultat de dl.istex.fr*.
- Cliquer sur Importer les données.
- Importer le modèle fourni, cliquer sur le menu en haut à droite *Modèle > Importer un modèle*.
- Publier votre site 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 à gauche. Consulter quelques ressources grâce à l'onglet *Recherche*.



## Étape 2

Importer le corpus dans Lodex.



◦ ◦ Combien y-a-t'il de documents publiés en 2009 dans la revue *Journal of Refugee Studies* ?

🔍 Vous pouvez saisir votre recherche ici ✕

Date de publication 2009 ✕ Titre de revue Journal of Refugee Studies ✕ Tout supprimer

<input type="checkbox"/>	exclure	
Titre de revue filtre		
	VALEUR	COMPTER ↓
<input type="checkbox"/>	PLOS ONE	188
<input checked="" type="checkbox"/>	Journal of Refugee Studies	180
<input type="checkbox"/>	International Journal of Refugee Law	134
<input type="checkbox"/>	Refugee Survey Quarterly	84
<input type="checkbox"/>	Journal of Immigrant and Minority Health	69
<input type="checkbox"/>	Social Science & Medicine	45
<input type="checkbox"/>	PLOS Medicine	29
<input type="checkbox"/>	Transcultural Psychiatry	24

18 ressources trouvées sur un total de 2036

☰ TRIER ASCENDANT ▾ 📄 EXPORTER ▾

**Representing Sahrawi Refugees' 'Educational Displacement' to Cuba: Self-sufficient Agents or Manipulated Victims in Conflict?** ▾

Mainstream accounts of refugee women and children have habitually portrayed their objects of study as 'generic' passive victims of war and/or famine. In stark contrast, however, since the early 1980s Sahrawi refugee women and children have been invoked as active agents constructing and maintaining their camps. In the first part of this article I explore the nature of a selection of mutually-reinforcing images produced by the Polisario Front (the Sahrawi refugees' 'representatives' and camp managers) for a European audience. Their distinctive content...

Journal of Refugee Studies

2009

**Contemporary Compulsory Dispersal and the Absence of Space for the Restoration of Trust** ▾

This paper investigates the issue of trust, or mistrust, specifically in relation to single adult asylum seekers and asylum seeker families compulsorily dispersed across England. It draws upon doctoral research on the social exclusion of asylum seekers as a result of dispersal and their separation from mainstream welfare provision due to the creation of the National Asylum Support Service (NASS) following the Immigration and Asylum Act 1999. Trust is an ambiguous term and four forms of trust are delineated to assist conceptualizing the experience of forc...

Journal of Refugee Studies

2009

**The Faint Footprint of Man: Representing Race, Place and Conservation on the Mozambique-South Africa Borderland** ▾

The integration of South Africa's existing Kruger National Park with Mozambique's newly-established Limpopo National Park signified an important milestone in the development of the Great Limpopo Transfrontier Park. Marketed as an example of the African Renaissance-in-action, this move towards a transnational 'mega-park' has promised an enlightened approach to conservation and regional economic development. For Mozambique, the LNP is a highly visible post-war initiative, attracting much needed investment to a remote and under-developed...

Journal of Refugee Studies

2009

# Étape 3.1

Extraction de mots-clés des résumés via le web service [Teeft](#).

- Aller dans *Données > Enrichissements* et cliquer sur *+ Ajouter*.
- Donner le nom *Mots-clés (WS)*, aller chercher l'url du web service **Teeft** dans le catalogue en cliquant sur le bouton vert à droite du champ *URL du web service*.



Sur quels types de champ peut-on lancer **Teeft** ? A quoi faut-il faire attention ?



# Étape 3.1

Extraction de mots-clés des résumés via le web service [Teef](#).

- Aller dans *Données > Enrichissements* et cliquer sur *+ Ajouter*.
- Donner le nom *Mots-clés (WS)*, aller chercher l'url du web service **Teef** dans le catalogue en cliquant sur le bouton vert à droite du champ *URL du web service*.



Sur quels types de champ peut-on lancer **Teef** ? A quoi faut-il faire attention ?

- Choisir *Résumé* dans la *Colonne de la source*, cliquer sur *Sauvegarder*. Cliquer enfin sur *Lancer*.

## Étape 3.2

Création du graphique associé à Teeft.

- **Dans Lodex, avant de faire un graphique, il est nécessaire de déclarer la colonne comme une ressource.** 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)*. Cliquer sur *Sauvegarder*.

## Étape 3.2

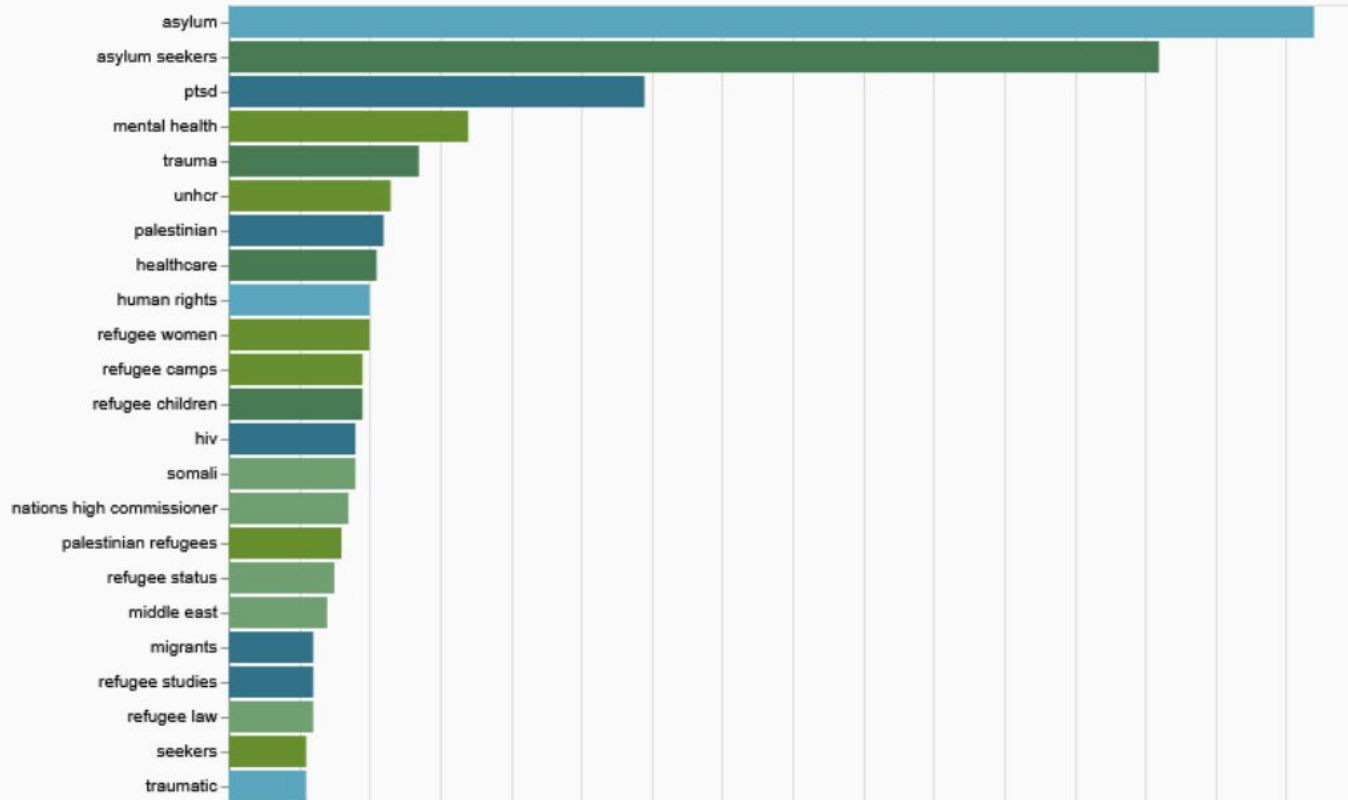
Création du graphique associé à Teeft.

- Pour créer le graphique : aller dans *Affichage* > *graphiques* cliquer sur + *Nouveau champ*, nommer le graphique en renseignant *Mots-clés les plus représentés dans les résumés* dans le champ *Étiquette*. Choisir la routine *distinct-by* puis choisir le champ sur lequel la routine va s'appliquer grâce au menu déroulant (soit le champ *Mots-clés*). Enfin, dans *Affichage*, choisir le format *Diagramme en barres* en filtrant les résultats : dans *Paramètres des Données* mettre valeur minimum à afficher à 5 et choisir le tri *Descendant*. Cliquer sur *Sauvegarder*.



Quels sont les mots-clés les plus représentés dans le corpus ?

## Mots-clés les plus représentés dans les résumés ⚙️



## Étape 4 : Création d'une carte en fonction des pays mentionnés dans les articles

# Étape 4.1 : Extraction des entités nommées Unitex

- Certains enrichissements sont déjà effectués sur l'entièreté de la base Istex. C'est le cas des noms de lieux renseignés dans la colonne *Entités nommées (Unitex)* avec les autres entités nommées.

Entités nommées (...)	Types d'enrichisse...	Score qualité
{"date":["year 1","1951"]	["grobidFulltext","multi	7.389
undefined	["grobidFulltext","multi	6.819
undefined	["nb","teft"]	9.827
{"date":["In the year 20	["grobidFulltext","multi	8.007
undefined	["nb","teft"]	10
undefined	["nb","teft"]	10
{"date":["2010-06-01"]	["grobidFulltext","multi	8.667
{"date":["1980s"],"orgh	["grobidFulltext","multi	8
undefined	["nb","teft"]	10
{"date":["the end of ye	["grobidFulltext","multi	7.599

```

4 : string "Mandy Savage"
]
  "placeName" : [ 7 items
    0 : string "Utrecht"
    1 : string "Europe"
    2 : string "America"
    3 : string "Amsterdam"
    4 : string "Arnhem"
    5 : string "England"
    6 : string "Netherlands"
  ]
  "ref_url" : [ 2 items
    0 : string "http://www.quantitativesk
    1 : string "http://www.r-project.org"
  ]
  "geogName" : [ 2 items
    0 : string "Netherlands Region"
    1 : string "Eastern Region"
  ]
}

```

## Étape 4.1 : Extraction des entités nommées Unitex

- Pour créer une colonne contenant uniquement ces noms de lieux, aller dans *Données > Enrichissements* et cliquer sur *+ Ajouter*. Donner le nom *Entités nommées de géographie*, cliquer sur *Mode avancé* et coller le code ci-après, cliquer sur *Sauvegarder*. Cliquer enfin sur *Lancer*.

```
[assign]
path = value
value = get("value.Entités nommées (Unitex).placeName").split(",").get(0)
```

*Remarque : le 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](#).*

## Étape 4.2 : Création de la carte

- Pour construire la carte, nous avons besoin du code pays à 3 lettres, retourné par le web service *Associer un terme au vocabulaire Pays et Subdivisions*. Aller dans *Données > Enrichissements > + Ajouter*, puis donner le nom *Pays et subdivisions (WS)*, en utilisant le web service *Associer un terme au vocabulaire Pays et Subdivisions* sur la colonne nouvellement créée *Entités nommées de géographie*. Cliquer sur *Sauvegarder*. Cliquer enfin sur *Lancer*.

Nom

Pays et subdivisions (WS) ▶ LANCER

0 erreurs RÉESSAYER

Statut : En cours [VOIR LES LOGS](#)

Mode avancé

URL du web service

https://loterre-resolvers.services.istex.fr/v1/9SD/identify 📄

Colonne de la source

Entités nommées de géographie ▼ Sous-chemin ▼

SUPPRIMER ANNULER SAUVEGARDER



# Étape 4.2 : Création de la carte

Accueil > Web-services > Associer un terme au vocabulaire Pays et Subdivisions

## Associer un terme au vocabulaire Pays et Subdivisions

### Description

### Utilisation

**Niveau d'utilisation :** Débutant

**Niveau de validation :** Expérimental

### Objectif

Le web service repère des termes dans le vocabulaire [Pays et Subdivisions de Loterre](#) pour récupérer le concept associé. Par exemple, il récupère la forme normalisée du nom d'un pays à partir de ses différentes variations orthographiques.

**Ressource :** La plateforme Loterre propose des vocabulaires ou terminologies pour différents domaines scientifiques. Autour d'un concept, Loterre propose plusieurs verbalisations et différents liens vers des bases de références. Le vocabulaire propose pour chaque pays et région française des concepts regroupant informations géographiques, variantes syntaxiques, acronymes, et formes normalisées.

# Étape 4.2 : Création de la carte

- Dans *Affichage > Ressource principale*, créer une ressource *Code pays* à partir de l'enrichissement *Pays et subdivisions (WS)*. Après avoir sélectionné la colonne, cliquer sur *Ajouter une opération* pour appliquer une transformation GET, renseigner *cartographyCode* dans le *path*. Cliquer sur *Sauvegarder*.

The screenshot displays a configuration interface for a resource named "Code Pays". A modal dialog titled "Ajouter une opération" is centered on the screen. Inside the dialog, a dropdown menu is set to "GET", and the "path" field contains the text "cartographyCode". Below the path field are two buttons: "ANNULER" and "CONFIRMER".

In the background, the main interface shows a table with the following columns:

- Etiquette: Code Pays
- Colonne(s) existante(s): Pays et subdivisions (WS)
- Opérations de transformation: TOUT SUPPRIMER

On the right side, a preview of the data is shown under the heading "Code Pays":

```
"n/a"
{"cartographyCode":"LUX","about":"http://data.lotterre.fr/ark:/67375/9SD-..."}
{"about":"http://data.lotterre.fr/ark:/67375/9SD-..."}
{"cartographyCode":"NZL","about":"http://data.lotterre.fr/ark:/67375/9SD-..."}
"n/a"
"n/a"
"n/a"
"n/a"
"n/a"
```

A note at the bottom right reads: "\* aperçu de la valeur avec les opérations de transformation appliquées".

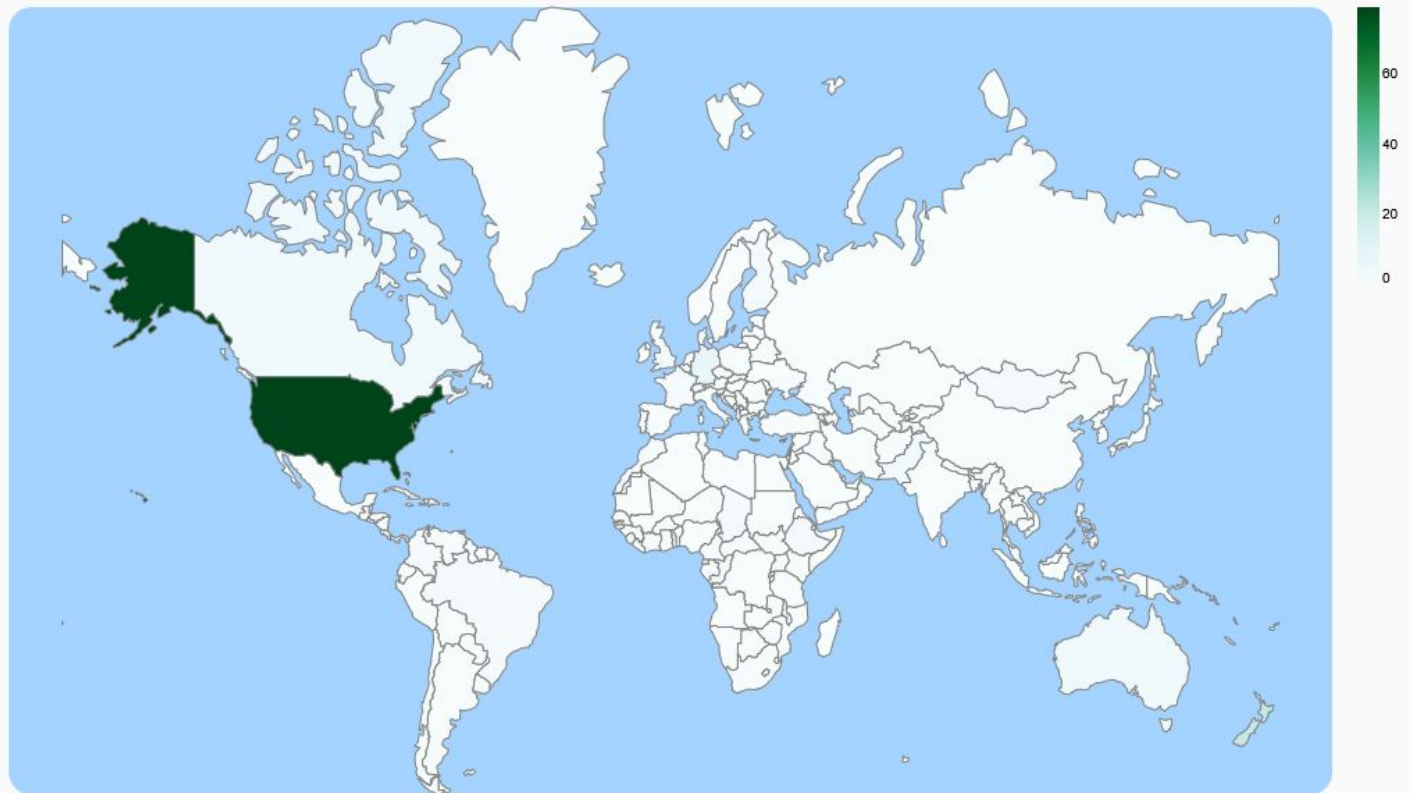
## Étape 4.2 : Création de la carte

- Pour créer le graphique : aller dans *Affichage > graphiques* cliquer sur + *Nouveau champ*, nommer le graphique en renseignant *Cartographie des pays cités* dans le champ *Étiquette*. Choisir la routine *distinct-by* puis choisir le champ sur lequel la routine va s'appliquer *Code pays*. Enfin, dans *Affichage*, choisir le format *Cartographie*. Cliquer sur *Sauvegarder*.



Quels sont les pays les plus mentionnés dans le corpus ?

## Cartographie des pays cités ⚙



# Étape 5

Extraction des thématiques du corpus.

- Dans *Données* > *Précalculs* cliquer sur *+ Ajouter*, lancer le précalcul *lda-segment* sur la colonne *Résumé*.



Quelle est la différence entre un web service et un précalcul ?

WS = prend en entrée “une seule ligne”

précalcul = prend en entrée tout un corpus

# Étape 5

Extraction des thématiques du corpus.

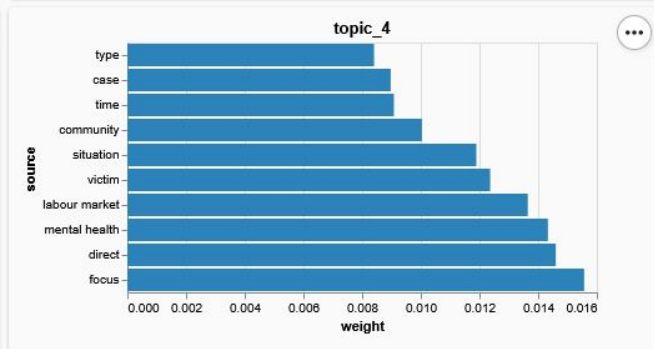
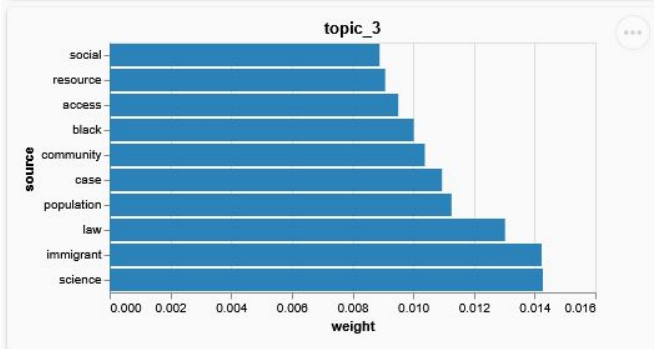
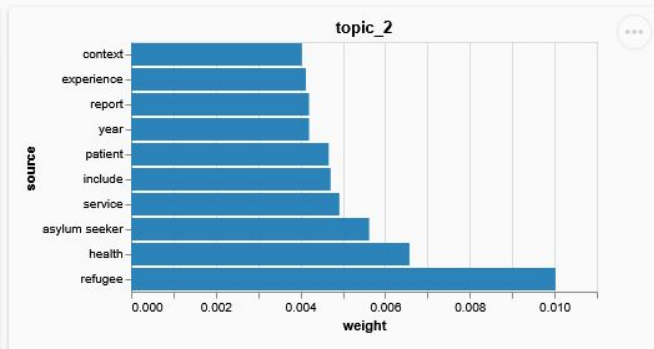
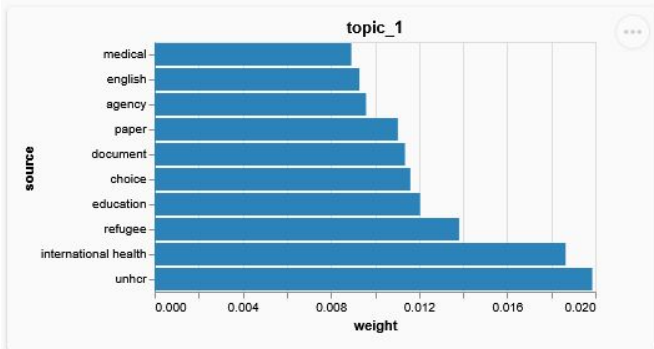
- **Les précalculs peuvent directement être utilisés pour faire des graphiques.** Aller dans *Affichage > graphiques* cliquer sur *+ Nouveau champ*, puis créer le graphique *Thématiques extraites du corpus* à partir du précalcul *lda-segment*. Choisir la routine *segments-precomputed-nofilter*. Enfin, dans *Affichage*, choisir le format *diagramme à barres groupées*.



◦◦. Combien de thématiques sont retournées ?

# Étape 5

LDA ⚙️



## Étape 6 : Pour aller plus loin...

L'objectif ici est d'augmenter l'exhaustivité de la carte. Nous pouvons utiliser un web service de détection d'entités nommées sur le corpus pour extraire des localisations.



En vous aidant d'[Istex TDM](#), quel web service vous semble le plus adapté ?

- Utiliser ce web service.
- Extraire les noms de lieux grâce à un nouvel enrichissement avancé en s'assurant de remplacer correctement "XXXXX" :

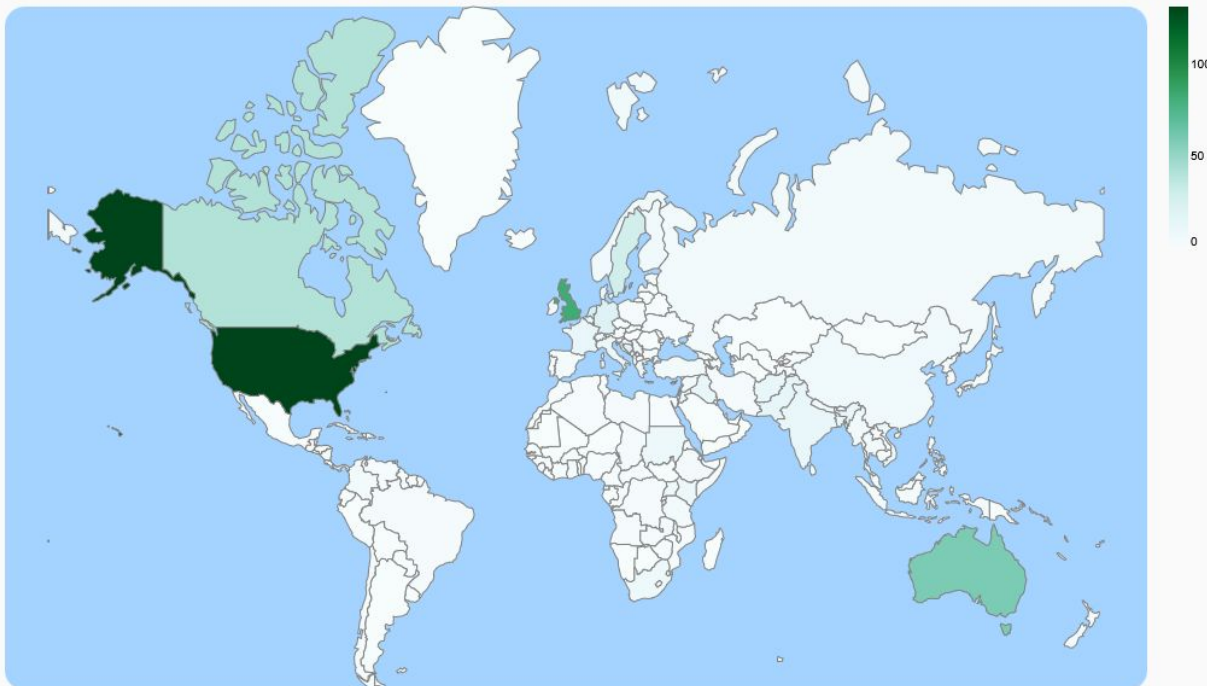
```
[assign]
path = value
value = get("value.XXXXX.LOC").split(",").get(0)
```

- Utiliser de nouveau le web service *Associer un terme au vocabulaire Pays et Subdivisions* sur les noms de lieux extraits à l'étape précédente (LOC). Créer une carte.



# Étape 6

Seconde cartographie ⚙



# Correction

Le TP corrigé est disponible en ligne :

<https://pantheon-sorbonne-14077.formation.lodex.fr/instance/correction-tp/>

Nom d'utilisateur : *correction*

Mot de passe : *TP2025*