



Usage des ressources ISTEX

Programme

1. Présentation générale d'ISTEX
2. Constitution d'un corpus par interrogation d'ISTEX

----- Une pause ! -----

3. Téléchargement du corpus avec ISTEX-DL
4. Exploration et valorisation du corpus avec l'outil LODEX
5. Liens utiles
6. Annexes

Documents utiles

Télécharger ce support et les fichiers utiles aux TP

<https://formation.lodex.fr>



1. *Enregistrer le zip contenant les documents de travail : clic droit, puis "enregistrer la cible du lien sous" → Bureau ou dossier créé*
2. *extraire les 5 fichiers : clic droit, puis "extraire tout" → Bureau ou dossier créé*

Index of /

[../Documents-atelier.zip](#)

04-Mar-2020 08:36

57956091

Introduction à ISTEX

1

ISTEX...

Initiative d'excellence en Information Scientifique et Technique
pour

Construire le socle de la bibliothèque scientifique numérique nationale



Construire le socle de la bibliothèque scientifique numérique nationale.



*est lancé le 30
avril 2012*

*est né d'une impulsion nationale, dans le
cadre du programme « Investissements
d'avenir » du Ministère de l'Enseignement
Supérieur, de la Recherche et de
l'Innovation (MESRI), et financé par
l'Agence Nationale de la Recherche (ANR)*

*est réservé aux abonnés de
l'Enseignement supérieur et de la
recherche française (ESR).*

*est porté par quatre partenaires : CNRS, ABES,
Couperin.org, l'Université de Lorraine pour la
Conférence des Présidents d'Université (CPU).*

[En savoir plus](#)

ISTEX : quels objectifs ?

- Acquisition massive et centralisée d'archives scientifiques
 - collections rétrospectives de la littérature scientifique
 - toutes disciplines
- Construction d'une plateforme nationale d'hébergement et de mise à disposition des données (Inist)

<https://www.istex.fr/>



La plateforme ISTEX : quels services ?

The screenshot displays the ISTEX platform interface. At the top, there is a navigation bar with a hamburger menu on the left and icons for a heart, a bell, and a grid on the right. The main content area features the ISTEX logo in the center. To the left of the logo, it states "23 millions de littérature scientifique" and "9 279 revues en accès libre". Below this, there is a search bar with the placeholder text "Testez ISTEX : indiquez un titre, des mots-clés". To the right of the logo, there is a section titled "de plus de 100 000 documents en accès libre de l'ESR". A large green arrow points from the right side of the interface towards the center, highlighting the services listed in the central panel.

Bouton	Scholar	Zotero
Télécharger	API	Harvester
SPARQL	data.istex.fr	Rechercher

La plateforme ISTEX : quels usages ?

Deux usages : l'un
classique, l'autre plus
avancé !

- Usage documentaire
- Usage en fouille de
textes (text mining)

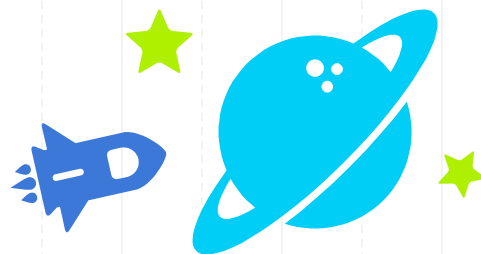


La fouille de textes désigne toute technique de **traitement automatisé** visant à **extraire des connaissances** d'un ensemble de textes sous forme numérique, produits par des humains pour des humains. Elle permet d'analyser parallèlement de **vastes quantités de données** selon un **critère de nouveauté ou de similarité**, et ainsi de dégager des **informations de haute qualité** difficiles à appréhender par la simple lecture cursive, telles que des **constantes**, des **tendances** et des **corrélations**.

(Sources : https://cnnumerique.fr/files/2017-10/CNNum_Fiche_TDM.pdf ;

https://fr.wikipedia.org/wiki/Fouille_de_textes ;

https://adbu.fr/competplug/uploads/2016/12/v9-Designed-Exec-Summary-ADBU-8pp_fra_final.pdf)



ISTEX

Quel contenu ? Quelques chiffres...





ISTEX aujourd'hui, c'est :

23 207 399

documents

27

corpus éditeurs

9 279

revues

346 863

monographies

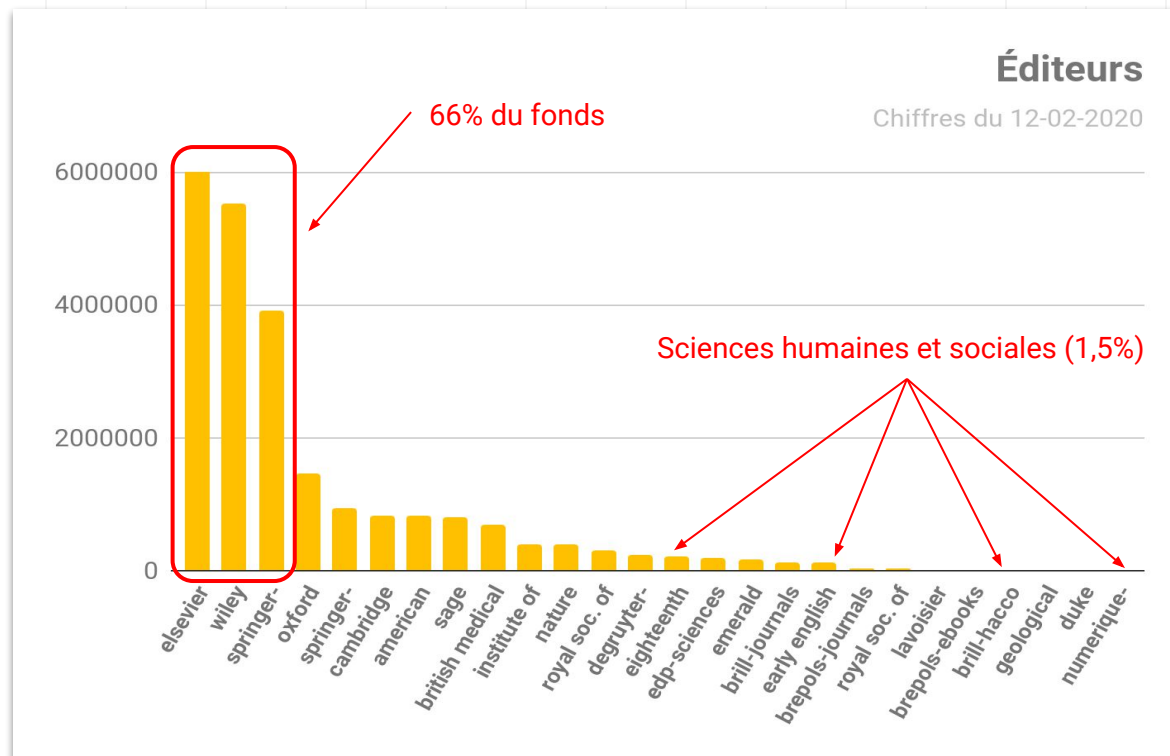
*Le fonds ISTE
en quelques
chiffres**

** Chiffres du 10-03-2020*



Le fonds ISTEEX > répartition par provenance

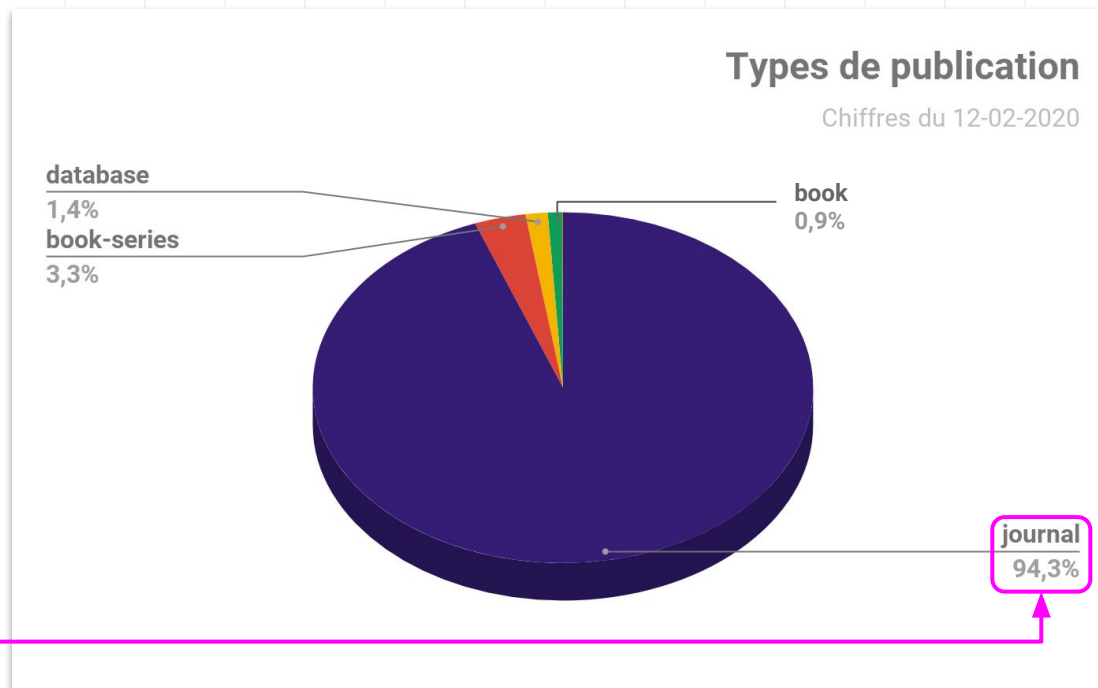
NB : Documents
provenant de 27 corpus
éditeurs mais chiffres
des graphiques datant
du 12/02/2020



Le fonds ISTEEX > répartition par types de publication

Prépondérance des documents issus de revues

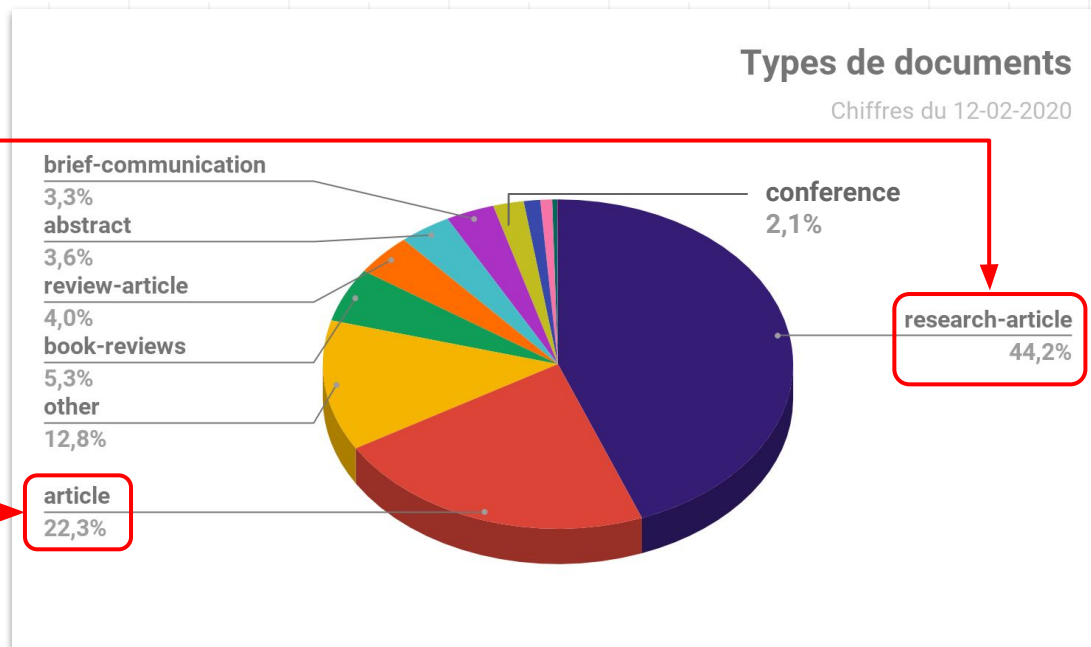
Définitions et chiffres sur data.istex.fr :
Types de publication



Le fonds ISTEEX > répartition par types de contenu

Plus de **66%** des documents sont des **articles**

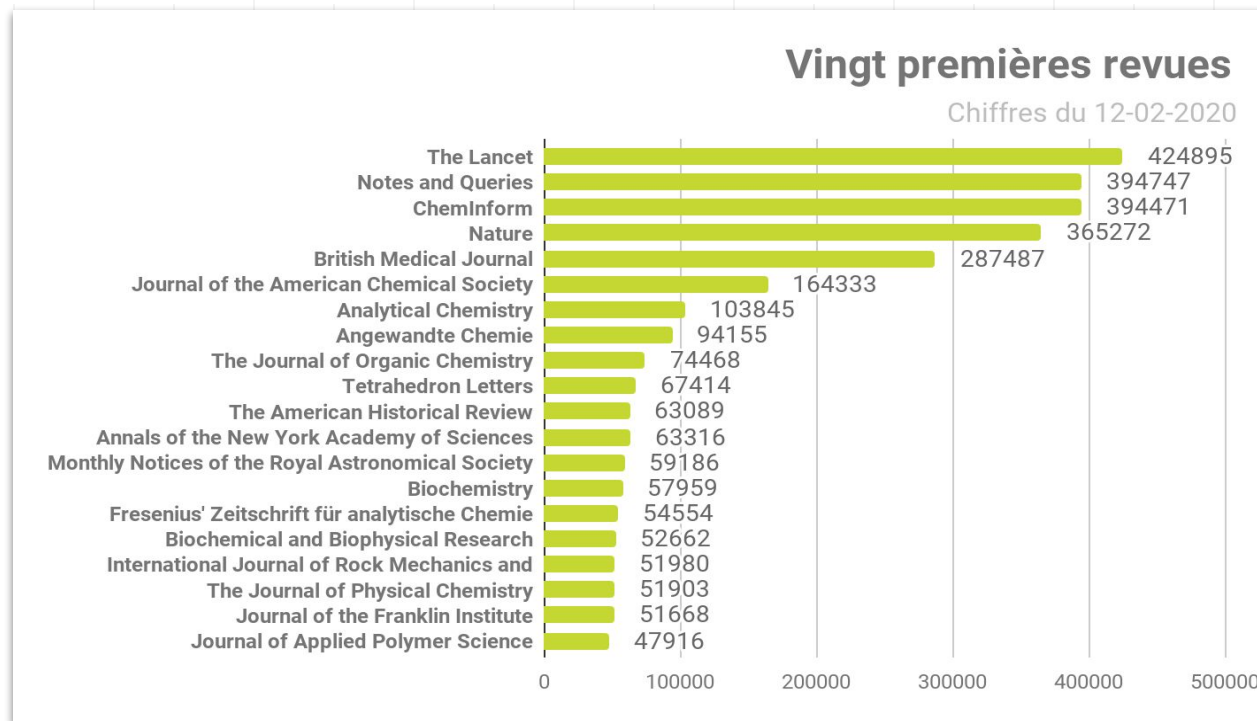
Définitions et chiffres sur
data.istex.fr :
Types de contenu



Le fonds ISTEEX > répartition par revues

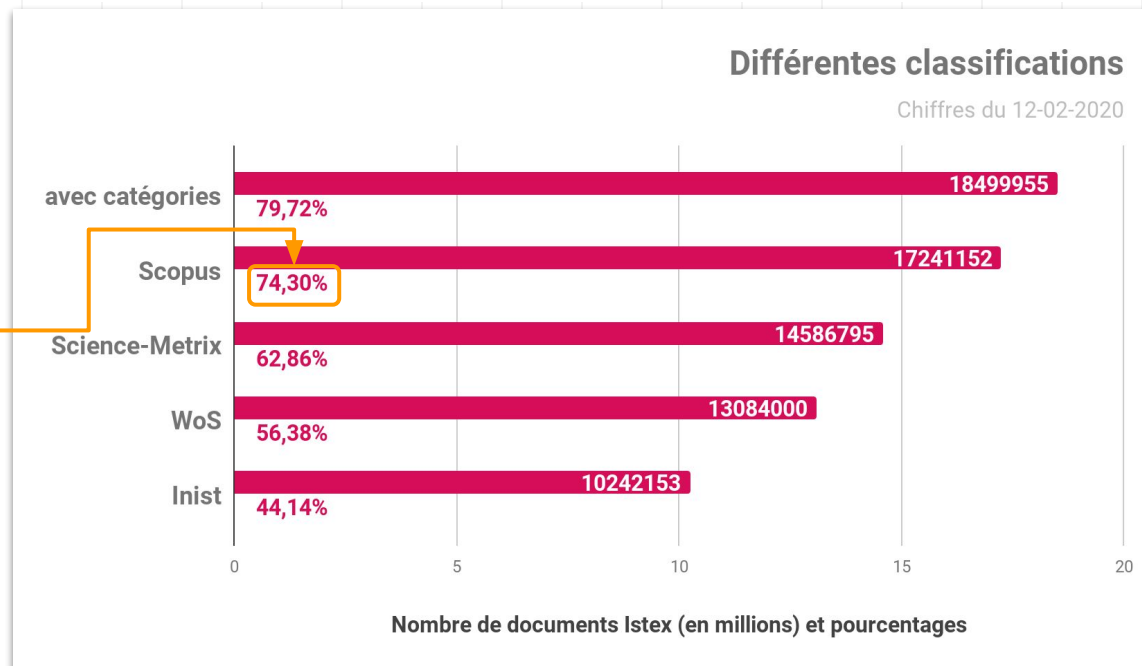
Dans le fonds de plus de **9 000** revues présentes dans ISTEEX :

liste des **20** revues les plus importantes en **nombre de documents**



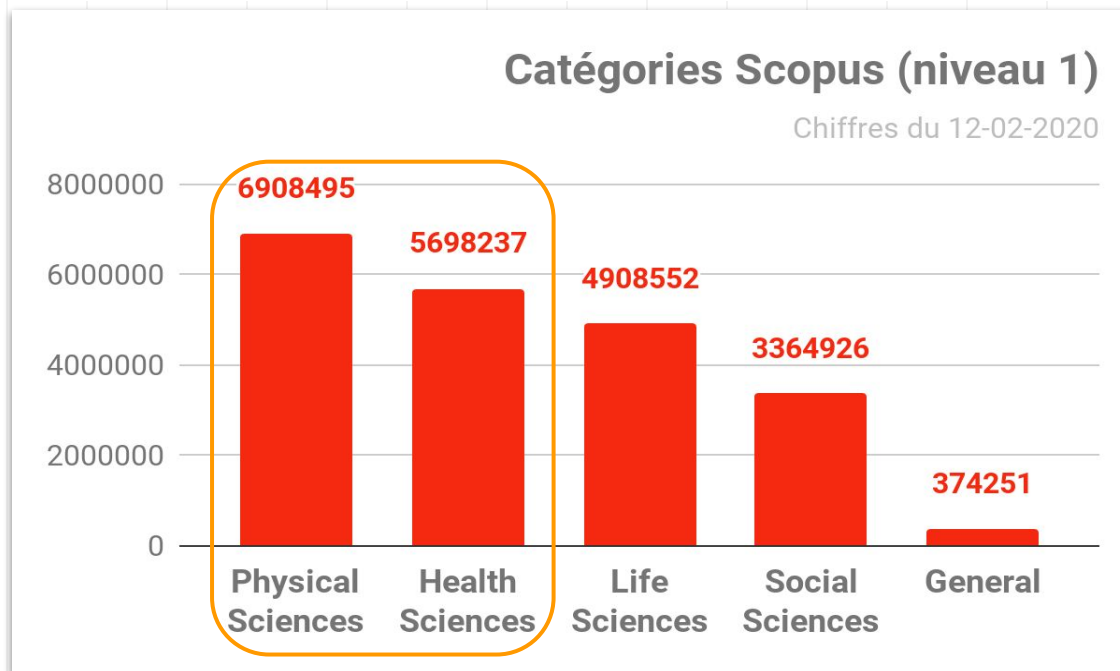
Le fonds ISTEEX > répartition par domaines scientifiques - 1

Scopus est la classification la plus représentative des documents Istex, sachant que plus de 20% des documents ne comportent aucune catégorie scientifique



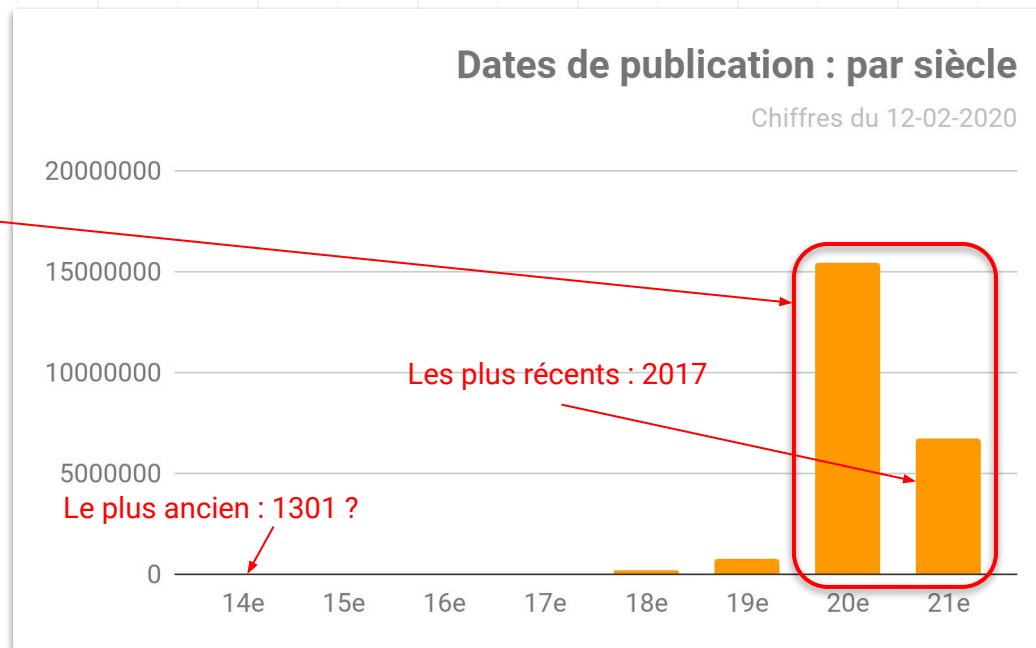
Le fonds ISTEEX > répartition par domaines scientifiques - 2

54% des documents Istex sont classés avec une catégorie Scopus relative aux sciences physiques ou de la santé



Le fonds ISTEEX > répartition par dates - 1

95% du fonds a été
publié aux 20^e et 21^e
siècles



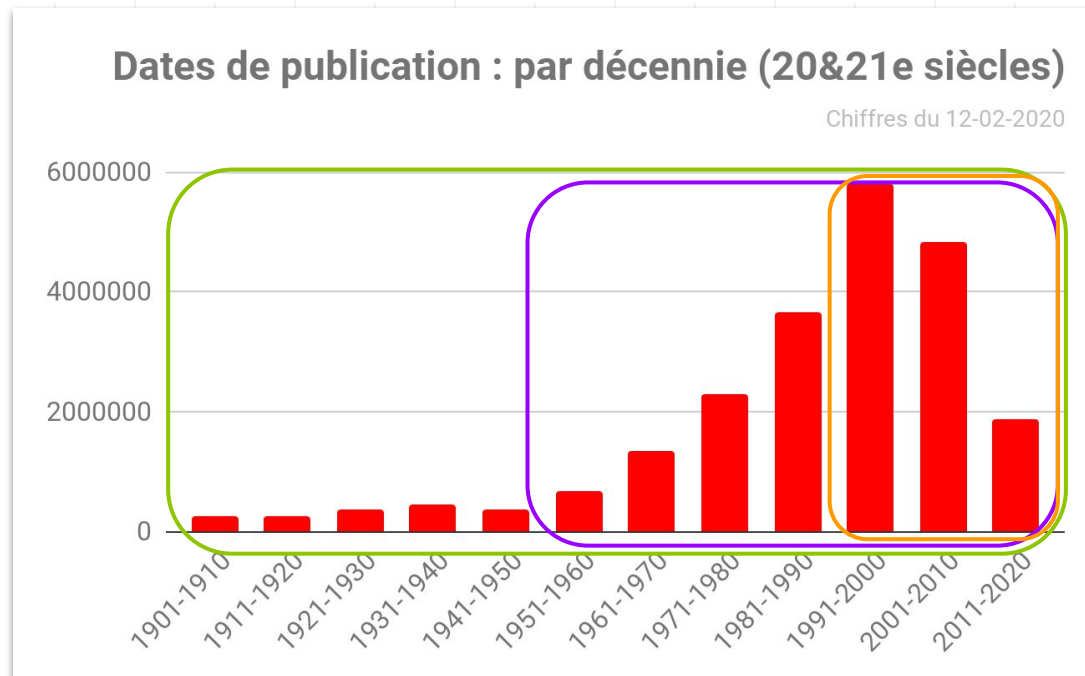
700 ans de publications

Le fonds ISTEK > répartition par dates - 2

95% des documents publiés
entre 1900 et aujourd'hui
(2017)

88% des documents publiés
depuis 1950

53% des documents publiés
sur les 30 dernières années



Le fonds ISTEK > répartition par langues - 1

53 langues !

Anglais
majoritaire

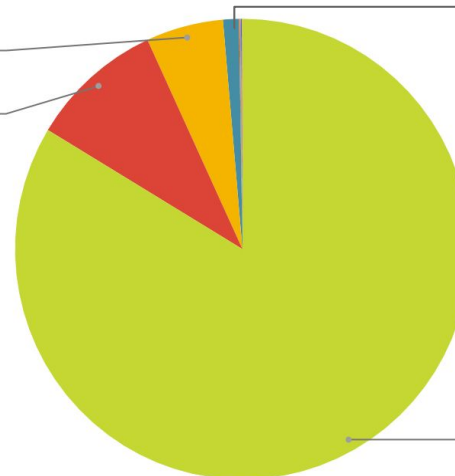
Information non
renseignée par les
éditeurs pour près
de 2,2 millions de
documents !

Langues majoritaires

Chiffres du 12-02-2020

allemand
5,5%
indéterminé
9,5%

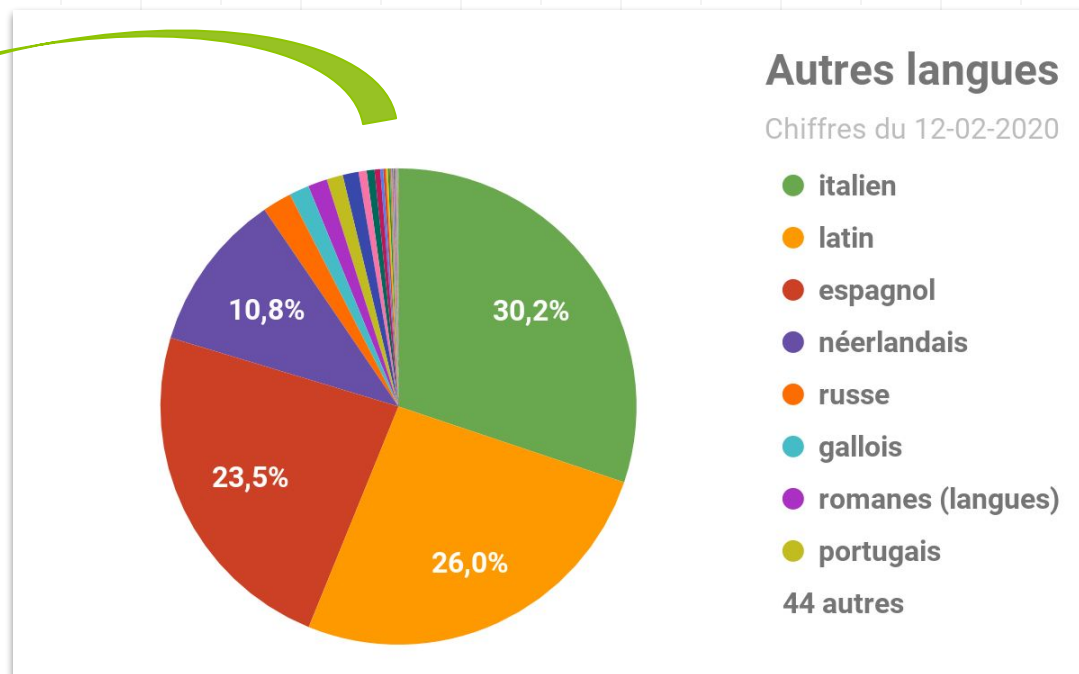
français
1,1%



anglais
83,7%

Le fonds ISTEK > répartition par langues - 2

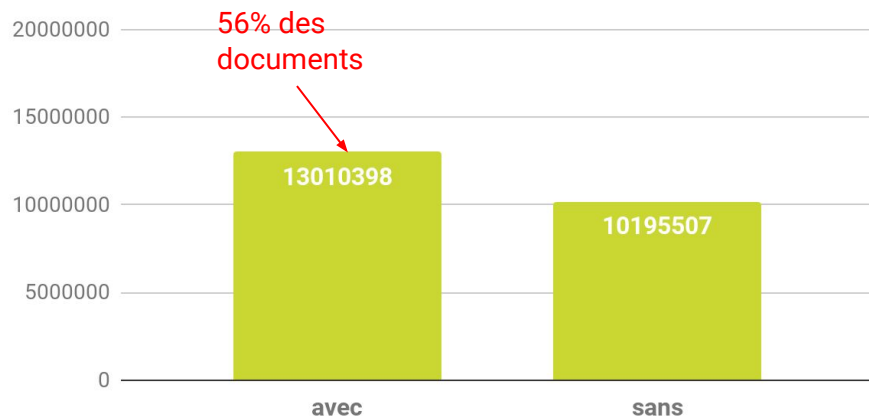
Les **49** autres langues concernent **0,3%** des documents restants (soit près de 57 000 documents)



Le fonds ISTEEX > répartition par éléments descriptifs du contenu

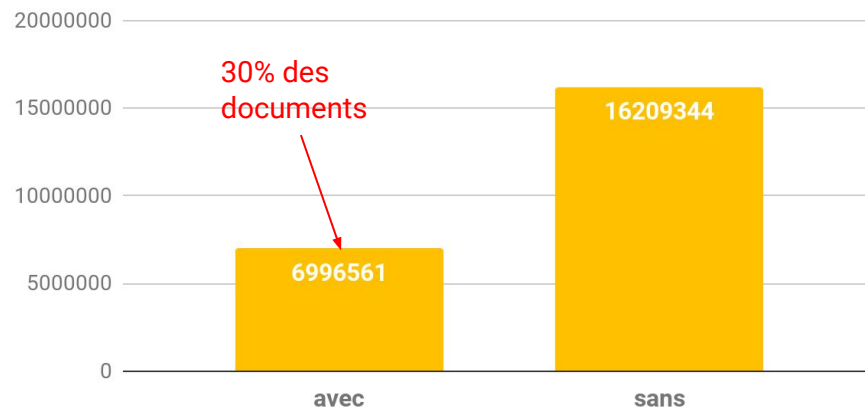
Présence de résumés

Chiffres du 12-02-2020



Présence de mots-clés d'auteur

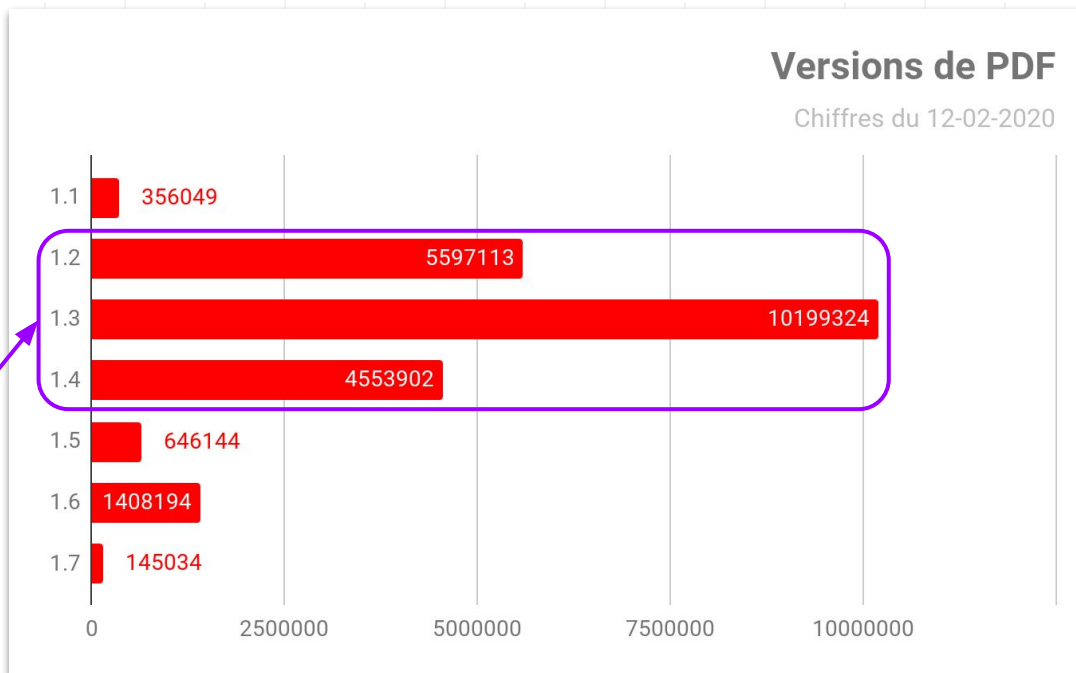
Chiffres du 12-02-2020



Le fonds ISTEK > répartition par indicateurs - 1

Plus la version de PDF est élevée, meilleure devrait être la qualité du document

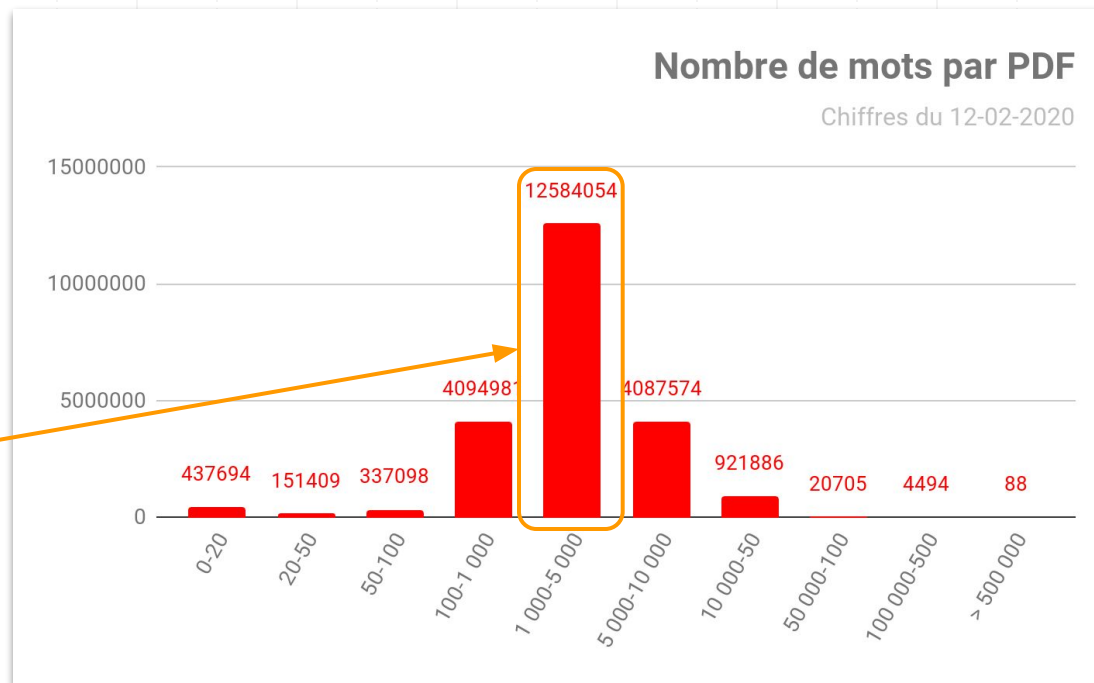
Près de 88% des documents existent dans une version de PDF comprise entre 1.2 et 1.4



Le fonds ISTEEX > répartition par indicateurs - 2

Un nombre de mots réduit peut indiquer un PDF image ou bien un court résumé

La majorité des documents PDF (54%) ont un nombre de mots compris entre 1 000 et 5 000

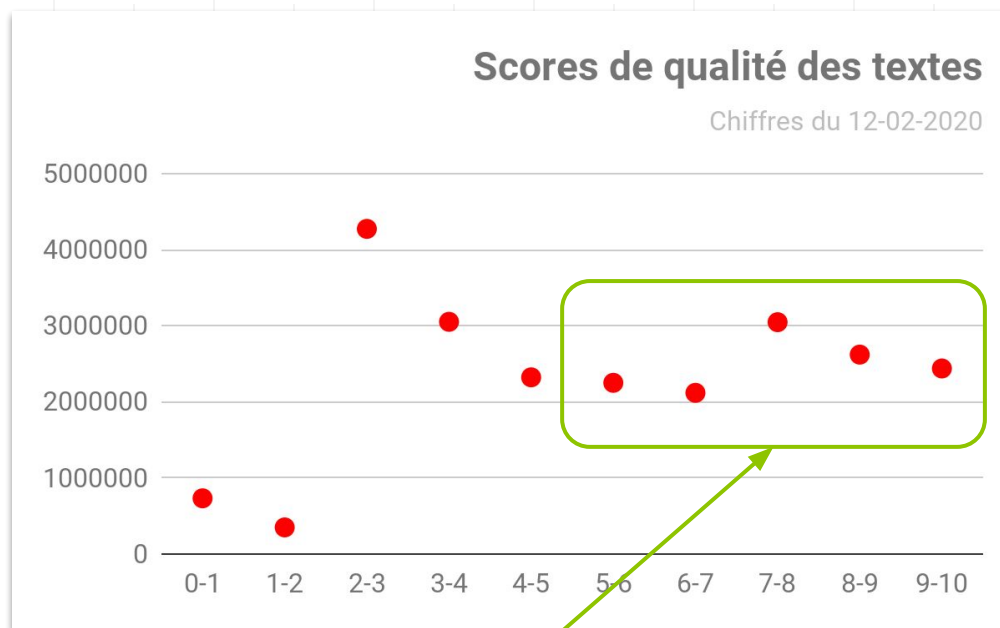


Le fonds ISTEK > répartition par indicateurs - 3

Le score de qualité est un nombre décimal compris entre 0 et 10.

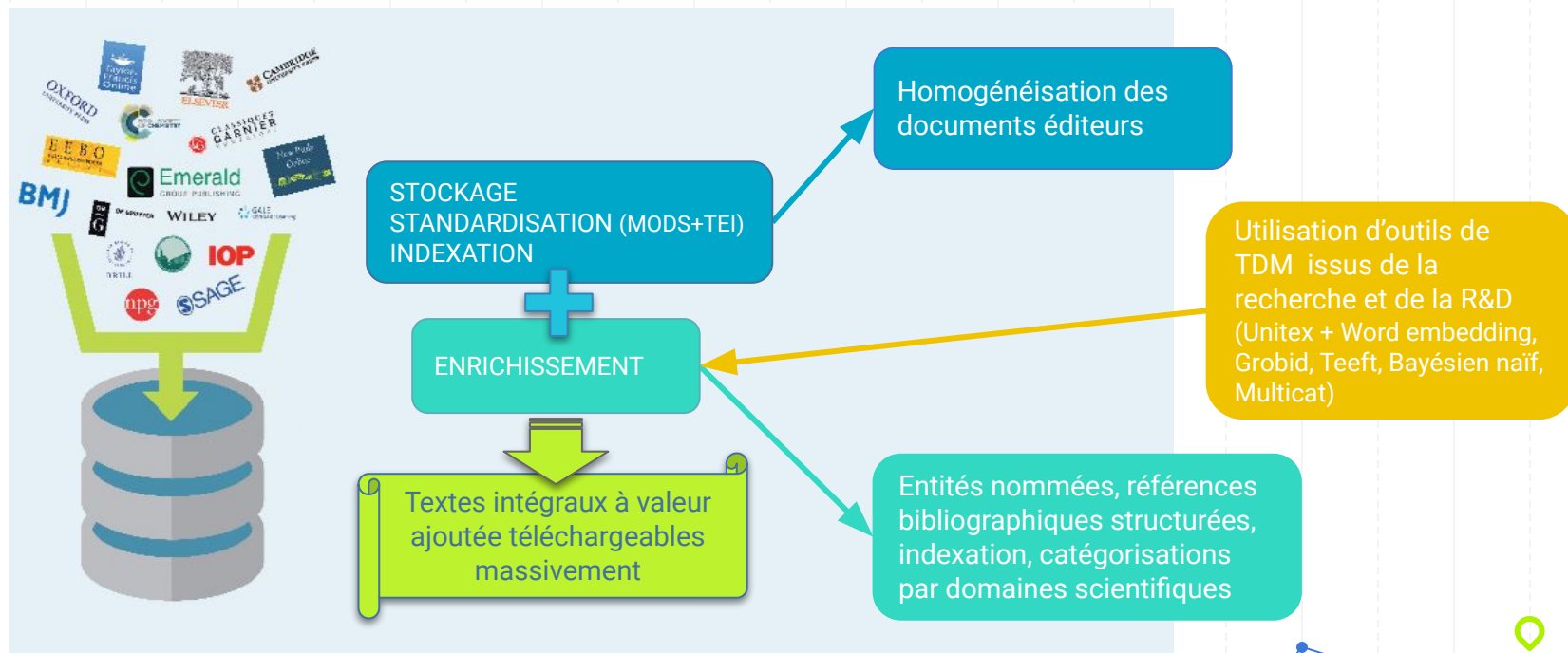
Son mode de calcul tient compte de la version du PDF, du nombre de mots du résumé, ainsi que du nombre de mots du texte intégral (PDF).

Voir la formule dans : Calcul du score de qualité



54 % des documents (12 475 415) ont un score supérieur à 5

La plateforme ISTE > la chaîne de traitements



La plateforme ISTEK : les enrichissements



Unitex

Détection de 9 types d'entités nommées (lieux géographiques, organismes financeurs, etc.)

enrichment-process.data.istex.fr/unitex

⇒ **68,6%** des docs



Bayésien naïf (nb)

Catégorisation de l'article par apprentissage avec Pascal/Francis

enrichment-process.data.istex.fr/nb

⇒ **43,4%** des docs



Grobid

Structuration des références bibliographiques

enrichment-process.data.istex.fr/grobid

⇒ **58,1%** des docs



Multicat

Catégorisation de la revue par appariement (WoS / Science Metrix / Scopus)

enrichment-process.data.istex.fr/wos

enrichment-process.data.istex.fr/science-metrix

enrichment-process.data.istex.fr/scopus

⇒ **75,2%** des docs



Teeft

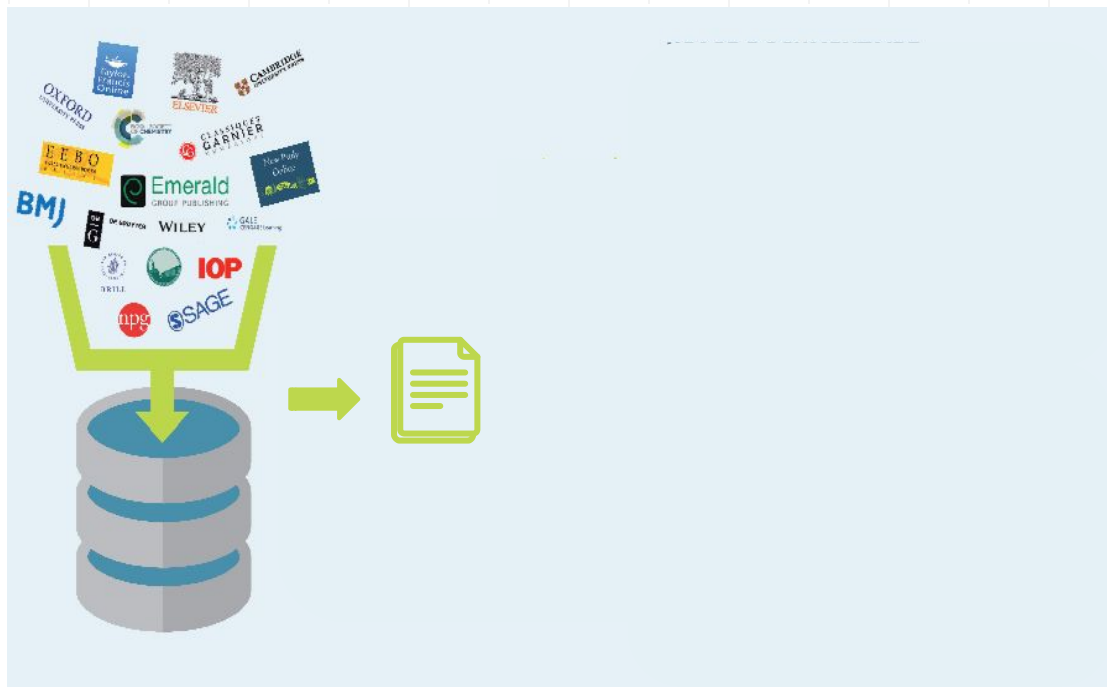
Indexation des articles par extraction de termes à partir du texte intégral en anglais

enrichment-process.data.istex.fr/teeft

⇒ **72,6%** des docs

Chiffres du 04-02-2020

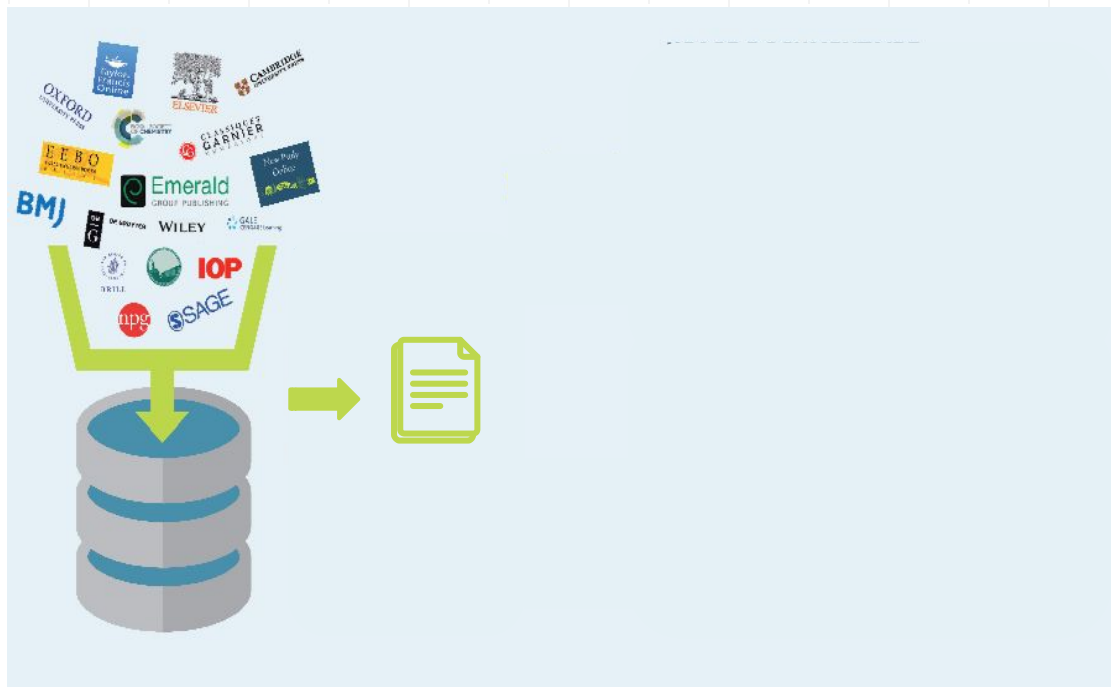
La plateforme ISTEK : deux usages



Exemples d'usage

- Accéder à un article
- Rechercher les articles d'un auteur
- Constituer une bibliographie sur un domaine
- Produire une synthèse documentaire

La plateforme ISTE : deux usages



Exemples d'usage TDM

- Aider au pilotage scientifique & au travail du chercheur
- Répondre à une question scientifique
- Usages hors Recherche

La plateforme ISTEK : exemples d'usage TDM*

Pilotage & Recherche

- Détection de sujets de recherche émergents
- Détection de plagiat
- Recommandation d'articles pour des néophytes ou non

** Ces cas d'usage sont documentés dans le fichier "Exemples-usage-TDM.pdf" présent dans le zip "Documents-atelier"*

Question scientifique

- Applications inattendues pour un ancien médicament
- Cartographie du génome humain
- Nouveaux matériaux thermoélectriques
- Traitement médiatique des "Gilets Jaunes"
- Portrait-robot d'espèces de poissons sauvages domesticables

Autres usages

- Marketing ciblé : détection d'intention d'achats et personnalisation des réponses
- Aide au diagnostic médical
- Prévention de la cybercriminalité
- Gestion des risques de fuite dans les réseaux d'assainissement
- Traduction automatique
- Assistant personnel

La plateforme ISTEK : exemples d'usage TDM à l'Inist

Réaliser des analyses scientifiques

- Analyses syntaxiques et terminologiques
- Evolution temporelle de thématiques

Mettre au point ou tester des outils

- **Entity-Fishing/IRC3** : détection de noms d'espèces animales et végétales
- **Unitex** : détection de définitions de concepts scientifiques
- **TermSuite** : détection de termes et de leurs variantes
- **TermSuite/Teeft + Neurodoc** : constitution et analyse de corpus
- **Topic Modeling, Iramuteq** : exploration thématique

Évaluer la performance d'outils

- **Unitex** : Corpus gold annoté manuellement en entités nommées pour évaluer les résultats de détection de l'outil

La plateforme ISTEEX : exemples d'usage TDM à l'Inist

Corpus réalisés à l'Inist :

- Sur des **domaines** et des **thématiques** variés :
 - Astrophysique, orthophonie, botanique, zoologie, géosciences, sciences et techniques alimentaires
 - Zone polaire arctique, vieillissement
- De petits et de gros corpus :
 - De **34** à **52 238** documents

Accessibles sur
data.istex.fr

Interrogation d'ISTEX

2

Interroger ISTE > Comprendre la structuration du réservoir

Le réservoir ISTE contient des **objets documentaires**

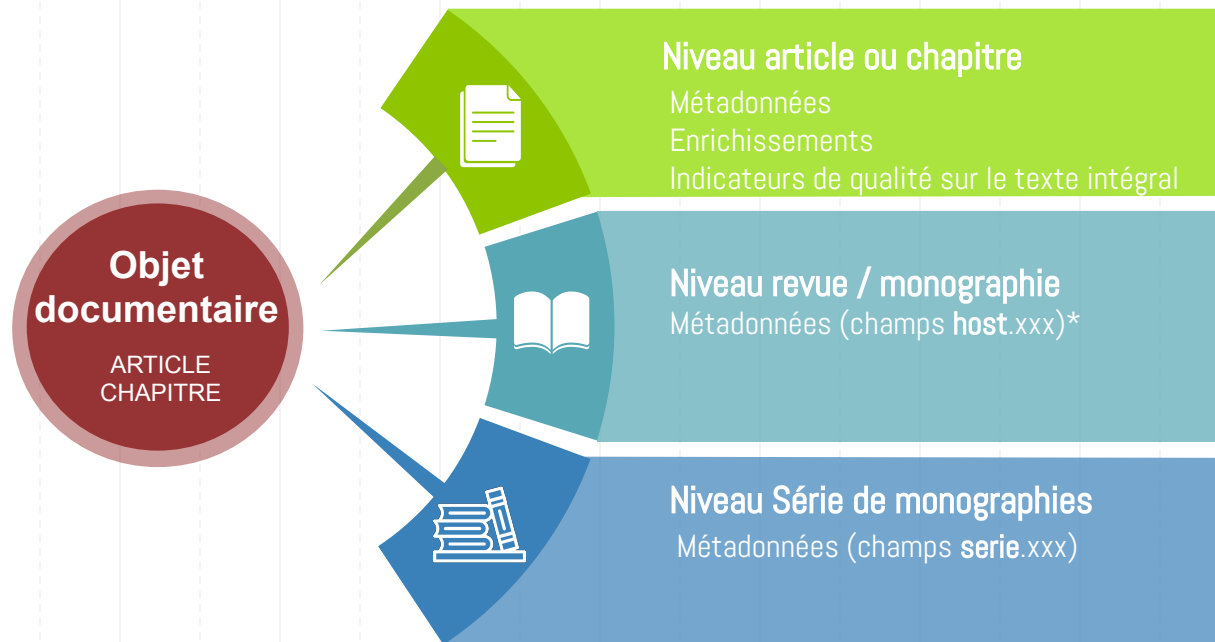
Ces objets documentaires sont :

- Soit des **articles** (de revue)
- Soit des **chapitres** (de monographie, d'e-book, de monographies en série)

Différents niveaux de granularité

L'interrogation peut se faire sur des informations concernant :

- Le niveau article ou chapitre
- Le niveau monographie
- Le niveau série de monographies



* Ces informations seront identiques pour **tous les articles d'une même revue** ou pour **tous les chapitres d'une même monographie**

Où interroger ISTEK ?

Le démonstrateur est une interface à **vocation pédagogique** branchée sur l'API ISTEK qui permet de :

- construire sa requête (en mode simple ou en mode avancé)
- visualiser et filtrer les résultats

<https://demo.istex.fr>



Interroger ISTEX > le démonstrateur

Les formats disponibles pour le texte intégral, les métadonnées décrivant le document et les annexes/couvertures

Bienvenue sur le **démonstrateur ISTEX**

[En savoir plus](#)

Résultats : 23205905 (639 ms) 1/ 2320591

[Mn12O12(OMe)2(O2CPh)16(H2O)2]2- Single-Molecule Magnets and Other Manganese Compou...

A new synthetic procedure has been developed in Mn cluster chemistry involving reductive aggregation of permanganate (MnO₄⁻) ions in MeOH in the presence of benzoic acid, and the first products from its use are described. The reductive aggregation of NBun4MnO₄ in MeOH/benzoic acid gave the new 4MnIV, 8MnIII anion...

Fulltext

Metadata

Annexes



Enrichments

[acs](#) [research-article](#)

[Inorganic Chemistry](#)

[artic/67375/TPS-5XN6BMF-5-Z](#)

[Score : 10](#)

[Mots : 9910](#)

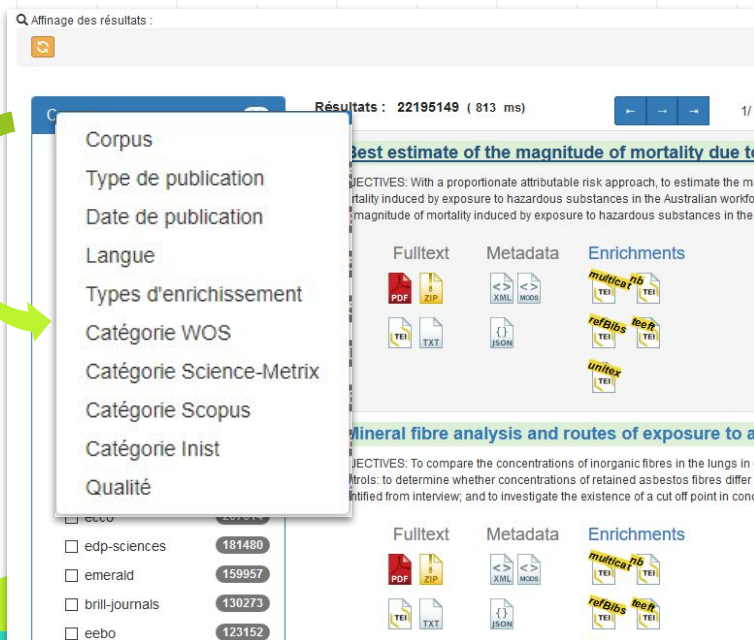
[Publication : 2005](#)

Accès rapide à différentes infos bibliographiques du document

Les différents types d'enrichissements disponibles en TEI

Interroger ISTE > le démonstrateur

Facettes pré-définies dans l'interface



- donne une vision synthétique du corpus
- permet de filtrer les résultats de la requête
- mais possibilités limitées - exploratoire

Interroger ISTE > le démonstrateur

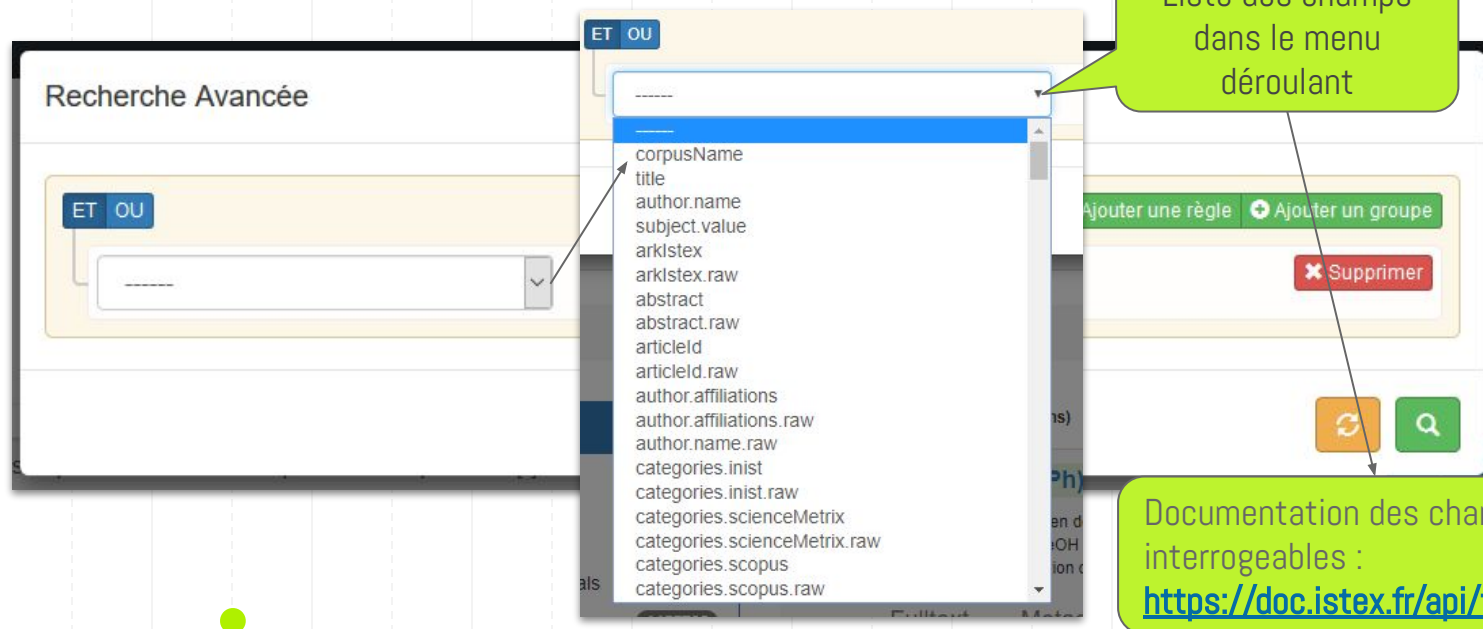
Bienvenue sur le démonstrateur ISTE

[En savoir plus](#)

Titre ou mot clef



Interroger ISTE > le démonstrateur



The screenshot shows the 'Recherche Avancée' (Advanced Search) interface. It features a search bar with a dropdown menu open, displaying a list of fields. A green callout bubble points to this dropdown menu, stating 'Liste des champs dans le menu déroulant'. Another green callout bubble points to a 'Documentation des champs interrogeables' link, stating 'Documentation des champs interrogeables : <https://doc.istex.fr/api/fields/>'. The interface also includes buttons for 'ET OU', 'Ajouter une règle', 'Ajouter un groupe', and 'Supprimer', along with search and refresh icons.

Recherche Avancée

ET OU

corpusName
title
author.name
subject.value
arklstex
arklstex.raw
abstract
abstract.raw
articleId
articleId.raw
author.affiliations
author.affiliations.raw
author.name.raw
categories.inist
categories.inist.raw
categories.scienceMetrix
categories.scienceMetrix.raw
categories.scopus
categories.scopus.raw

Ajouter une règle + Ajouter un groupe

Supprimer

Documentation des champs interrogeables : <https://doc.istex.fr/api/fields/>

Interroger ISTE > la recherche basique

Petit exercice pour démarrer :

- On souhaite rechercher les documents possédant le terme **"polar"**

Bienvenue sur le **démonstrateur ISTE**

[En savoir plus](#)

 [Options](#)

[Recherche avancée](#)



Recherche sur :

- Les **métadonnées**
- Le texte **intégral**
- Les **références** bibliographiques
- Les **enrichissements**



Résultats : + de 1 000 000 docs !!!

Cas d'usage

Objectif pédagogique :

Écrire une équation, testée pas à pas, utilisant un certain nombre d'opérateurs, d'astuces et de syntaxes, associés à une recherche sur un certain nombre de champs utiles pour délimiter un corpus pertinent et de taille raisonnable



"Zoologiste spécialisé en milieu marin au département "Adaptations du vivant" du MNHN, je m'intéresse à l'adaptation au changement global en zone polaire"

Interroger ISTE > Cas d'usage

Application TDM

- Outil de reconnaissance d'**entités nommées** scientifiques
- **Détecter** les **espèces animales**
- Dans le texte intégral

Thématique

- zoologie : **espèces animales** de type **poisson** ou **mollusque**
- zone géographique : zone **polaire** ou **subpolaire arctique**

Contraintes liées à l'application

- L'outil traite les documents **en anglais**
- Les documents devront être en format **PDF**
- Attention : l'outil ne traite pas les **PDF image**



Rédiger sa requête dans un bloc-note type "NotePad".
Ou copier successivement les exemples de "TP.txt",
et coller les requêtes dans le démonstrateur

Interroger ISTEX > Principes de base

1. je recherche les termes suivants
 - fish, mollusc
 - arctic, subarctic, Svalbard, Barents sea
2. dans les champs
 - titre
 - résumé
 - mots-clés d'auteurs
3. je limite mon corpus aux documents
 - langue de publication = anglais

Notions abordées

Opérateurs booléens

Parenthésage

Recherche sur champs

Recherche insensible à la casse

Expressions multitermes / mots-composés

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie

fish **OR** mollusc

Explications :

- **OR** cumule les résultats associés à "fish" et "mollusc"
- Les mots sont recherchés sur tout le document
- Si pas d'opérateur utilisé, l'opérateur par défaut OR s'applique



*Les opérateurs doivent s'écrire **en MAJUSCULES***

Résultats (10 mars 2020) : 800 717 docs

Interroger ISTE > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie

(fish OR mollusc) AND

(arctic OR subarctic)

Explications :

- OR cumule les résultats associés à "fish" et "mollusc"
- Les mots sont recherchés sur tout le document
- Si pas d'opérateur utilisé, l'opérateur par défaut OR s'applique
- AND + parenthèses permet d'associer 2 critères, ici termes zoologiques et termes géographiques

Résultats (10 mars 2020) : 50 614 docs

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie
contenant **majuscule**

(fish OR mollusc) AND

(arctic OR subarctic OR svalbard)

Explications :

- OR cumule les résultats associés à "fish" et "mollusc"
- Les mots sont recherchés sur tout le document
- Si pas d'opérateur utilisé, l'opérateur par défaut OR s'applique
- AND + parenthèses permet d'associer 2 critères, ici termes zoologiques et termes géographiques
- insensibilité à la casse
 - "svalbard" cherche aussi "Svalbard"

Résultats (10 mars 2020) : 51 100 docs

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule, expression multiterme**

(fish OR mollusc) AND

(arctic OR subarctic OR svalbard OR
"barents sea")

Résultats (10 mars 2020) : 52 885 docs

Explications :

- OR cumule les résultats associés à "fish" et "mollusc"
- Les mots sont recherchés sur tout le document
- Si pas d'opérateur utilisé, l'opérateur par défaut OR s'applique
- AND + parenthèses permet d'associer 2 critères, ici termes zoologiques et termes géographiques
- insensibilité à la casse
 - "svalbard" cherche aussi "Svalbard"
- utilisation d'une expression multiterme entre guillemets
 - barents sea = 1 346 725 docs
 - "barents sea" = 12 752 docs

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule, expression multiterme, mot-composé**

(fish OR mollusc) AND

(arctic OR subarctic OR "sub-arctic"
OR svalbard OR "barents sea")

Résultats (10 mars 2020) : 52 885 docs

Explications :

- **OR** cumule les résultats associés à "fish" et "mollusc"
- Les mots sont recherchés sur tout le document
- Si pas d'opérateur utilisé, l'opérateur par défaut OR s'applique
- **AND + parenthèses** permet d'associer 2 critères, ici termes zoologiques et termes géographiques
- **insensibilité à la casse**
 - "svalbard" cherche aussi "Svalbard"
- utilisation d'une **expression multiterme** entre guillemets
 - **barents sea** = 1 347 835 docs
 - **"barents sea"** = 12 778 docs
- utilisation d'un **mot-composé** entre guillemets
 - **sub-arctic** = 2 116 617 docs
 - **"sub-arctic"** = 7 808 docs

Interroger ISTEX > Principes de base

Equation étape 1

Recherche sur des termes de zoologie et de géographie contenant **majuscule, expression multiterme, mot-composé**

(fish OR mollusc) AND

(arctic OR subarctic OR "sub-arctic"
OR svalbard OR "barents sea")



Langues autres que anglais : **sensibilité aux diacritiques**
Ex : "mammifère" ne cherche pas "mammifere"

Résultats (10 mars 2020) : 52 885 docs

Explications :

- OR cumule les résultats associés à "fish" et "mollusc"
- Les mots sont recherchés sur tout le document
- Si pas d'opérateur utilisé, l'opérateur par défaut OR s'applique
- AND + parenthèses permet d'associer 2 critères, ici termes zoologiques et termes géographiques
- insensibilité à la casse
 - "svalbard" cherche aussi "Svalbard"
- utilisation d'une expression multiterme entre guillemets
 - barents sea = 1 347 835 docs
 - "barents sea" = 12 778 docs
- utilisation d'un mot-composé entre guillemets
 - sub-arctic = 2 116 617 docs
 - "sub-arctic" = 7 808 docs

Plus de détails : [opérateurs](#) / [astuces](#)

Interroger ISTEX > Principes de base

Equation étape 2

Recherche restreinte sur certains champs

```
title:((fish OR mollusc) AND (arctic OR  
subarctic OR "sub-arctic" OR svalbard OR  
"barents sea"))
```

```
OR abstract:(idem)
```

```
OR subject.value:(idem)
```

Résultats (10 mars 2020) : 52 885 docs ➤ 1 243 docs

Explications :

- Recherche restreinte sur les champs :
 - titre de l'article : "title"
 - résumé : "abstract"
 - mots-clés d'auteur : "subject.value"
- Les noms de champs sont introduits par :
- Ajout de parenthèses pour regrouper les valeurs recherchées dans un même champ
- Pas de factorisation des noms de champs. Il faut répéter les termes pour chaque champ interrogé

Plus de détails : [exemples de contenus](#) / [recherche sur champs](#)

Interroger ISTE > Principes de base

Equation étape 3

Restriction aux documents publiés **en anglais**

(idem) AND **language:eng**

Explications :

- Ajout de **parenthèses** à l'ensemble de l'équation car on ajoute un critère qui porte sur toute l'équation précédente
- Recherche sur le champ : **language**

Liste des [codes langues](#)

Résultats (10 mars 2020) : 1 243 docs ➤ 1 133 docs

Interroger ISTE > Principes de base

Équation complète à cette étape

```
(title:((fish OR mollusc) AND (arctic OR subarctic OR svalbard OR "sub-arctic" OR "barents sea"))  
OR abstract:((fish OR mollusc) AND (arctic OR subarctic OR svalbard OR "sub-arctic" OR "barents sea"))  
OR subject.value:((fish OR mollusc) AND (arctic OR subarctic OR svalbard OR "sub-arctic" OR "barents sea")))  
AND language:eng
```

Résultats (10 mars 2020) : 1 133 docs

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

4. je limite mon corpus aux documents

- qui ne sont pas des PDF image

5. je cherche à ajouter un maximum de variantes pertinentes aux termes suivants

- fish, mollusc
- svalbard

Les variantes peuvent être de plusieurs natures :

- formes au **pluriel**
- formes **non accentuées**
- variantes **orthographiques**
- formes **composées**



Notions abordées

Interrogation sur des intervalles de nombres

Utilisation de troncatures

Utilisation d'expressions régulières

Interroger ISTE > Niveau Intermédiaire - affiner les valeurs

Equation étape 4

Un petit test pour commencer : Exploration de la facette "Qualité" dans le démonstrateur

Qualité ▾

Score

Entre : 2.434 à 9.934

Nombre de mots

Entre : 0 à 174600


Status of Zoobenthos and Fish Populations in Subarctic Rivers of the Northernmost Finland:...

Abstract: In 1990, a monitoring programme was initiated to survey the status of benthic invertebrate communities and fish populations over a wide range of subarctic rivers in northernmost Finnish Lapland (68°15'–70°N). A special emphasis was placed on detecting possible effects of acidification through sulfur...

Fulltext






Metadata





Enrichments




[springer-journals](#)
[research-article](#)
[ark:/67375/VQC-GQ4X0XS9-2](#)
 Score : 4.116
 Mots : 16
 Publication : 2001

1 - Réglage du curseur à 500

3 - Ouverture du PDF
et sélection du texte

2 - Vérification du
nombre de mots

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 4

Recherche restreinte sur certains champs avec des **intervalles de valeurs**

`qualityIndicators.pdfWordCount: [500 TO 10000]`

`qualityIndicators.pdfVersion: [1.2 TO *]`

Explications :

Recherche sur des intervalles de valeurs à l'aide de crochets []

[500 TO 10000] ← valeurs limites inférieure et supérieure

[1.2 TO *] ← valeur limite supérieure infinie



TO s'écrit obligatoirement en MAJUSCULES

Plus de détails : [intervalles](#)

Interroger ISTE > Niveau Intermédiaire - affiner les valeurs

Equation étape 5

Recherche sur des formes au pluriel et des formes non accentuées à l'aide de **troncatures**

fish* OR **mollus***

Explications :

- ? remplace **1** caractère
- * remplace **0 à n** caractère(s)

mollus* = mollusque(s) mollusca mollusc(s) mollusk(s) molluscan...

fish* = fish fishes fishing fishery fisheries fisherman fishermen...
et Tully-Fisher ! ⚠ **variante hors domaine**

Plus de détails : [troncatures](#)

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 5 (suite)

Recherche plus ciblée sur des formes au pluriel à l'aide d'expressions régulières

`/fish(es)?/`



sensibilité à la casse :

aucune majuscule entre les délimiteurs //

Explications :

La troncature par * ne convient pas ici car ramène des variantes hors domaine

Les expressions régulières permettent de rendre compte de toutes les possibilités d'écriture d'un terme en ciblant des variantes spécifiques

fish OR fishes = `/fish(es)?/`

terminaison "es"
optionnelle

Plus de détails : [expressions régulières](#)

Interroger ISTEX > Niveau Intermédiaire - affiner les valeurs

Equation étape 5 (suite)

On peut aller beaucoup plus loin avec la recherche sur des termes avec variantes orthographiques

Exemple : Spitzberg, île principale de l'archipel du Svalbard, s'écrit avec 4 orthographes possibles

svalbard OR /spit[sz]berg(en)?/

Explications :

Spit^sberg OR Spitz^zberg OR Spit^sberg^{en} OR Spitz^zberg^{en} =

/spit[sz]berg(en)?/

caractère au choix
entre ^s et ^z

terminaison "^{en}"
optionnelle

Plus de détails : [expressions régulières](#)

Interroger ISTE > Niveau Intermédiaire - affiner les valeurs

Équation complète à cette étape

```
(title:( (/fish(es)?/ OR mollus*) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea"))  
OR abstract:( (/fish(es)?/ OR mollus*) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea"))  
OR subject.value:( (/fish(es)?/ OR mollus*) AND (arctic OR subarctic OR "sub-arctic" OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea"))) )  
AND language:(eng)  
AND qualityIndicators.pdfWordCount:[500 TO 10000]
```

Résultats (10 mars 2020): 1 133 ➤ 1 192 docs

Interroger ISTEX > pour les plus avancés

6. je cherche à ajouter des variantes supplémentaires aux termes suivants :

- subarctic, barents sea

Notions abordées

Recherche floue

Recherche de proximité

Interroger ISTE > pour les plus avancés

Equation étape 6

Recherche sur des variantes d'écriture du terme "subarctic" à l'aide de la **recherche floue** :

"sub-arctic" OR **subarctic~1**



- permet de **recupérer** des articles **pertinents** bien que contenant des erreurs (numérisation)
- à utiliser **avec prudence** après **vérification** des résultats

Explications :

en fonction du nombre entier indiqué après le tilde, compris entre **0 et 2**, on recherche un terme ayant **au maximum 2 caractères de différence** (en plus, en moins ou dissemblables).

- subarctic~0 récupère : subarctic
- subarctic~1 récupère : subarctic + subarctic, subarctic, subarctic, subarctica
- subarctic~2 récupère : subarctic, subarctic, subarctic, subarctic, subarctica + subarctica **mais aussi** subarctic, subarctic 

NB : subarctic~ (sans chiffre précisé) équivaut à subarctic~2

Plus de détails : [recherche floue](#)

Interroger ISTE > pour les plus avancés

Equation étape 6 (suite)

Recherche sur des synonymes de "barents sea" à l'aide d'**opérateurs de proximité**

"barents sea" OR "barents seas"~2

Explications :

en fonction du nombre entier indiqué après le tilde, compris entre **0 et n**, on recherche une expression comportant deux termes **plus ou moins distants** (distance de **n termes**).

"barents seas"~2 récupère :

- Barents-Kara Seas
- Barents and Kara seas
- Barents, and Kara seas



Le moteur de recherche ne tient pas compte de l'ordre des termes et recherche les termes dans les 2 sens

NB : pas de troncature possible

Plus de détails : [recherche de proximité](#)

Interroger ISTE > pour les plus avancés

Équation complète

```
(title:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2))  
  
OR abstract:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2))  
  
OR subject.value:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2)))  
  
AND language:eng  
  
AND qualityIndicators.pdfWordCount:[500 TO 10000]
```

Résultats (10 mars 2020): 1 192 ➤ 1 197 docs

Interroger ISTEEX > Et après ?

Téléchargement du corpus avec ISTEEX-DL



Exploration de l'interface

Réutilisation de la requête constituée

Sélection du format de fichier en fonction du cas d'usage

Exploration du corpus dans LODEX



Import du fichier dans Lodox

Modélisation à l'aide du modèle prédéfini

Exploration des différents graphiques



Téléchargement de corpus

ISTEX-DL
& Autres outils

3

“

***ISTEX-DL ou ISTEX-DownLoad :
“télécharger un corpus ISTEX en
quelques clics”***

ISTEX-DL : Objectif

Interface web conviviale de type formulaire pour **extraire un corpus ISTEX** en **3 étapes** et récupérer son corpus sur son poste de travail dans un **format compressé (zip)**...avec un **minimum** de connaissance informatique !

<https://dl.istex.fr>



TROIS ÉTAPES,...

construction
d'une équation

sélection des
fichiers

téléchargement
du corpus

1. Requête

Explicitiez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique ⓘ

Recherche par ARK ⓘ

ark:/67375/OT8-JMF4G14B-2
ark:/67375/OT8-RNCBH0VZ-8

Nombre de caractères restants ⓘ : 67 000 ✓

Choisir le nombre de documents souhaités ⓘ : 50000

Choisir les documents classés ⓘ :

Par pertinence ⓘ Aléatoirement ⓘ

2. Formats et types de fichiers

Créez votre sélection en cochant ou décochant les cases ci-dessous :

☐ Texte intégral ⓘ

☐ PDF

☐ TEI

☐ TXT

☐ ZIP

☐ TIFF

☐ Métadonnées ⓘ

☐ JSON

☐ XML

☐ MODS

☐ Annexes ⓘ

☐ Couvertures ⓘ

☐ Enrichissements ⓘ

☐ multicat

☐ nb

☐ refBibs

☐ teeft

☐ unitex

3. Télécharger



...CINQ BOUTONS

On gomme...

Réinitialiser

On ne perd pas
sa requête en
cours...

Récupérer

On partage son
corpus...

On teste...

Exemples

Partager

Historique

On peut rejouer le
passé...



Exemples



Réinitialiser



Récupérer



Partager



Historique

Deux modes de recherche

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique

Recherche par ARK 

ark:/67375/0T8-JMF4G14B-2
ark:/67375/0T8-RNCBH0VZ-8

Nombre de caractères restants  : 67 000 

Choisir le nombre de documents souhaités  : 

Choisir les documents classés  :
☒ Par pertinence ☐ Aléatoirement

ISTEX-DL : étape 1

Recherche par identifiants des documents

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

ark:/67375/0T8-JMF4G14B-2
ark:/67375/0T8-RNCBH0VZ-8

ARK (équivalent du DOI)
attribué à chaque
document ISTEEX

Nombre de caractères restants  : 67 000 

Choisir le nombre de documents souhaités  :  0 50000 100000

Choisir les documents classés  :

☒ Par pertinence ☐ Aléatoirement

ISTEX-DL : étape 1

Recherche par identifiants des documents

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

ark:/67375/0T8-JMF4G14B-2
ark:/67375/0T8-RNCBH0VZ-8

Nombre de caractères restants  : 67 000 

Choisir le nombre de documents souhaités  : 

Choisir les documents classés  :
☒ Par pertinence ☐ Aléatoirement

 Exemples

 Réinitialiser

 Récupérer

 Partager

 Historique

ISTEX-DL : étape 1

Recherche par identifiants des documents



ISTEX-DL : étape 1

ISTEX

Vous êtes membre

1. Requête

Explicitiez ci-dessous l'équation

Recherche classique

ark:/67375/0T8-JM

ark:/67375/0T8-RNCBH0VZ

Nombre de caractères restant

Limite du nombre de documents

Exemples de requêtes

Voici quelques exemples dont vous pouvez vous inspirer pour votre recherche. Cliquez sur l'une des loupes et la zone de requête sera remplie automatiquement par le contenu de l'exemple choisi. Cet échantillon illustre différentes façons d'interroger l'API Istex en utilisant :

- des données bibliographiques
- des données bibliographiques et des indicateurs de qualité
- des données bibliographiques et des troncatures sur des mots-clés
- des troncatures sur des mots-clés et des opérateurs booléens imbriqués
- des expressions régulières sur des mots-clés (I)
- des expressions régulières sur des mots-clés (II)
- des troncatures, de la recherche floue et des opérateurs de proximité
- des identifiants ISTEX de type ARK

Annuler

Recherche par identifiants des documents



ISTEX-DL : étape 1

1. Requête i

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

[Recherche classique i](#) [Recherche par ARK i](#)

ark:/67375/HXZ-3PZ5S1MB-7
ark:/67375/HXZ-XH6SRM66-7
ark:/67375/6H6-K00Q3697-8
ark:/67375/6H6-52QRK0X3-F
ark:/67375/HXZ-H5NSV5QF-M
ark:/67375/HXZ-J1BLQKH4-3
ark:/67375/HXZ-QVSXSCHW-P
ark:/67375/HXZ-RVC4D13J-W
ark:/67375/6H6-19QHR7H2-6
ark:/67375/HXZ-0F8518D4-K
ark:/67375/6H6-9Q9HLC0X-7
ark:/67375/6H6-QTZPZZXD-H

Exemples

Réinitialiser

Récupérer

Partager

Historique

Nombre de caractères restants i : 66 689 ✓

L'équation saisie correspond à 12 document(s)

Limite du nombre de documents souhaités i : 12 12 0

75

Recherche par identifiants des documents

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

brain AND language:fre

Nombre de caractères restants  : 67 000 

Choisir le nombre de documents souhaités  :

50000



Choisir les documents classés  :

☒ Par pertinence ☐ Aléatoirement

ISTEX-DL : étape 1

Longueur
équation :
limitation

(nombre &
couleur)

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit

Recherche classique 

Recherche par ARK 

brain AND language:fre

Nombre de caractères restants  : -2 893 

Longueur autorisée
équivalent à 15 pages de
texte ou 2 500 ARK

Coloration orange à
moins de 1 000
caractères de la limite

Choisir le nombre de documents souhaités  :

50000

0

Limite
dépassée

Choisir les documents classés  :

☒ Par pertinence ☐ Aléatoirement

ISTEX-DL : étape 1

Recherche par équation booléenne



ISTEX-DL : étape 1

1. Requête

Exemple de requête n°2

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique

Recherche par ARK

host.title.raw("Marine Biology") AND qualityIndicators.pdfVersion:[1.2 TO *] AND qualityIndicators.score:[3.5 TO *]

Nombre de caractères restants : 66 884

L'équation saisie correspond à 10 179 document(s)

Choisir le nombre de documents souhaités : 10179

Choisir les documents classés :

Par pertinence

Aléatoirement

78

Recherche par équation booléenne



ISTEX-DL : étape 1

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

host.title.raw:("Marine Biologie" AND qualityIndicators.pdfVersion:[1.2 TO *] AND qualityIndicators.score:[3.5 TO *]
AND publicationDate:

Nombre de caractères restants  : 66 863 

L'équation saisie correspond à 0 document

Choisir le nombre de documents souhaités  :   0

Choisir les documents classés  :

☒ Par pertinence ☐ Aléatoirement

Erreur de syntaxe dans votre requête 

Syntaxe de la requête incorrecte sur : undefined . Si vous utilisez un caractère & dans votre requête, veuillez le remplacer par son équivalent ASCII : %26. Allez sur <https://api.istex.fr/documentation/300-search.html> pour plus d'informations sur la syntaxe des requêtes.

Recherche par équation booléenne



ISTEX-DL : étape 1

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

host.title.raw("Marine Biology") AND qualityIndicators.pdfVersion:[1.2 TO *] AND qualityIndicators.score:[3.5 TO *]
AND publicationDate:[1980 TO *]

Nombre de caractères restants  : 66 852 ✓

L'équation saisie correspond à 7 941 document(s)

Choisir le nombre de documents souhaités  :

7941

Choisir les documents classés  :

☒ Par pertinence ☐ Aléatoirement

Nombre de documents : limitation

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

arctid

Nombre de caractères restants  : 66 994 

Pas d'extraction possible au-delà de 100 000 documents

L'équation saisie correspond à 183 595 document(s) 

Choisir le nombre de documents souhaités  :

50000



ISTEX-DL : étape 1

Nombre de documents : limitation

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le contenu de votre recherche

Recherche classique

Recherche par ARK

arctic AND qualityIndicators.pdfVersion{1.3 TO *}

Nombre de caractères restants : 66 950

L'équation saisie correspond à 70 981 document(s)

Choisir le nombre de documents souhaités : 50000

Choisir les documents classés :
☒ Par pertinence ☐ Aléatoirement

Modifier l'équation pour réduire le nombre de documents

ISTEX-DL : étape 1

Nombre de documents : sélection

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

arctic AND qualityIndicators.pdfVersion:{1.3 TO *}]

Nombre de caractères restants  : 66 950 

L'équation saisie correspond à 70 981 document(s)

Choisir le nombre de documents souhaités  :

50000

Choisir les documents classés  :

☒ Par pertinence ☐ Aléatoirement

Avec le sélecteur ou la réglette, adapter le choix du nombre de documents à la durée d'extraction souhaitée et à la capacité de stockage

ISTEX-DL : étape 1

Deux modes de tri

mode par défaut qui trie selon un score de pertinence des documents en réponse à l'équation

ISTEX-DL : étape 1

1. Requête i

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique i

Recherche par ARK i

brain AND language:fre

Nombre de caractères restants i : 67 000 ✓

Choisir les documents classés i :

☐ Par pertinence ☒ Aléatoirement

000

50000

100000

mode à choisir pour constituer un échantillon représentatif (si réduction du nombre de documents)

Recherche par équation booléenne

Cas d'usage

ISTEX-DL : étape 1

1. Requête

Explicitez ci-dessous l'équation ou la liste d'identifiants qui décrit le corpus souhaité :

Recherche classique 

Recherche par ARK 

```
(title:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 )) OR abstract:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 )) OR subject.value:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))) AND language:eng AND qualityIndicators.pdfWordCount:[500 TO 10000]
```

Nombre de caractères restants  : 66 460 

L'équation saisie correspond à 1 197 document(s)

Choisir le nombre de documents souhaités  : 1197



Sélection

2. Formats et types de fichiers

Créez votre sélection en cochant ou décochant les cases ci-dessous :

☐ Texte intégral

- ☐ PDF
- ☐ TEI
- ☐ TXT
- ☐ ZIP
- ☐ TIFF

☐ Métadonnées

- ☐ JSON
- ☐ XML
- ☐ MODS

☐ Annexes

☐ Couvertures

☐ Enrichissements

- ☐ multicat
- ☐ nb
- ☐ refBibs
- ☐ teeft
- ☐ unitex

Un exemple de document

formats & fichiers disponibles

Eight glacial cycles from an Antarctic ice core

The Antarctic Vostok ice core provided compelling evidence of the nature of climate, and of climate feedbacks, over the past 420,000 years. Marine records suggest that the amplitude of climate variability was smaller before that time, but such records are often poorly resolved. Moreover, it is not possible to infer the abundance of...

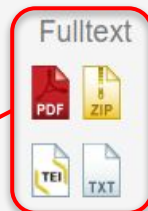
nature research-article

ark:/67375/GT4-NBDQR6K2-9

Score : 9.796

Mots : 6131

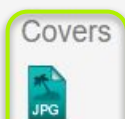
Publication : 2004



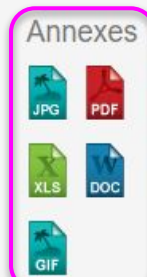
4 formats
(les plus courants)



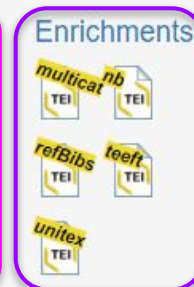
3 formats
toujours
présents



1 couverture
(format image)
disponible



5 annexes
disponibles
(avec des
formats
différents)



Les 5 types
d'enrichissement
présents dans Istex
(au format TEI)

ISTEX-DL : étape 2

Sélection des fichiers

2. Formats et types de fichiers

Créez votre sélection en cochant ou décochant les cases ci-dessous :

- ☐ Texte intégral 
- ☐ PDF
 - ☐ TEI
 - ☐ TXT
 - ☐ ZIP
 - ☐ TIFF

Le format OCR, s'il existe pour un document, sera préféré au TXT

- ☐ Métadonnées 
- ☐ JSON
 - ☐ XML
 - ☐ MODS

Liste de tous les formats possibles dans ISTE

- ☐ Annexes 
- ☐ Couvertures 

- ☐ Enrichissements 
- ☐ multicat
 - ☐ nb
 - ☐ refBibs
 - ☐ teeft
 - ☐ unitex

Extraction des types disponibles pour chaque document

ISTEX-DL : étape 2

Pas de détail des formats

Quels fichiers
seront
extraits ?

Aide des
couleurs



ISTEX-DL : étape 2

2. Formater les fichiers

Créez votre sélection en cochant ou décochant les cases ci-dessous :

☐ Texte intégral

- ☐ PDF
- ☐ TEI
- ☐ TXT
- ☐ ZIP
- ☐ TIFF

☒ Texte intégral

- ☒ PDF
- ☒ TEI
- ☒ TXT
- ☒ OCR
- ☒ ZIP
- ☒ TIFF

☒ Texte intégral

- ☐ PDF
- ☐ TEI
- ☒ TXT
- ☒ OCR
- ☐ ZIP
- ☐ TIFF

☐ Enrichir

- ☐ nb
- ☐ refBibs
- ☐ teeft
- ☐ unitex

Sélection /
désélection
automatique de
tous les formats
en cochant /
décochant le
niveau fichier

Indication
d'une
sélection
partielle

Quels fichiers choisir ?

Aide des infobulles



ISTEX-DL : étape

ISTEX

Taper pour rechercher

Documentation ISTEX

Usage documentaire d'ISTEX

Usage TDM d'ISTEX

Construction d'une requête

Extraction d'un corpus

Vérification et mise en forme des r...

Annexes

Liste des codes langues dans IS...

Liste des types de publication e...

Liste des catégories scientifique...

Liste des formats et types de ...

API ISTEX

Intégrer ISTEX dans un ENT

FAQ

Texte intégral

o PDF (Portable Document Format **Format de document portable*)

Il s'agit d'un format de description de pages pouvant contenir du texte, des dessins, des images et photographies (noir et blanc, couleur, 3D). C'est un format ouvert, évolutif et multiplateforme, issu de l'imprimerie, qui conserve la mise en page du document original. Il offre une sécurité permettant à l'auteur d'un document d'empêcher sa modification par des utilisateurs. Il a été créé par Adobe Systems, Inc. Le logiciel Adobe® Acrobat® Reader est nécessaire pour lire et imprimer un fichier PDF.

Les fichiers en format PDF dans ISTEX sont des fichiers originaux fournis par l'éditeur.

o ZIP

Il s'agit d'un format permettant l'archivage et la compression de fichiers. L'archivage est l'utilisation d'un seul fichier pour stocker plusieurs fichiers. La compression des fichiers permet de réduire leur taille. Compresser les fichiers permet de gagner du temps dans le chargement des données et de la place dans le stockage de celles-ci. Le logiciel de compression analyse le fichier et compresse les parties qui se répètent. Lors de la décompression, la forme originale du fichier est restaurée. On peut le comparer à la

Published with GitBook

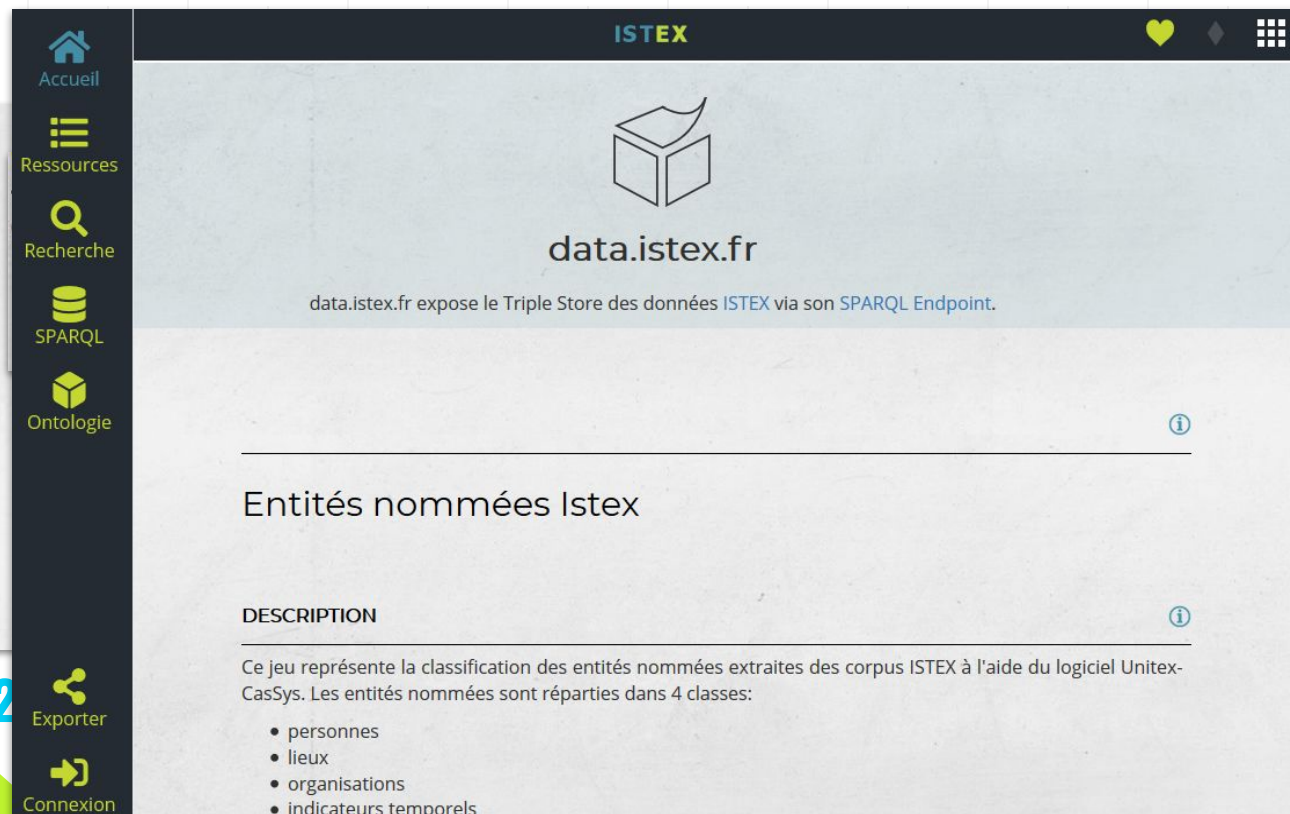
90

Quels fichiers
choisir ?

Aide des
infobulles



ISTEX-DL : étape 2



The screenshot shows the web interface of data.istex.fr. On the left is a dark sidebar with navigation links: Accueil (home icon), Ressources (list icon), Recherche (magnifying glass icon), SPARQL (cylinder icon), and Ontologie (cube icon). At the bottom of the sidebar are 'Exporter' (share icon) and 'Connexion' (arrow icon). The main content area has a header with the ISTEX logo, a heart icon, and a grid icon. Below the header is a large box with the text 'data.istex.fr' and 'data.istex.fr expose le Triple Store des données ISTEX via son SPARQL Endpoint.' Below this is a section titled 'Entités nommées Istex' with an information icon. Underneath is a 'DESCRIPTION' section with an information icon. The description text reads: 'Ce jeu représente la classification des entités nommées extraites des corpus ISTEX à l'aide du logiciel Unitex-CasSys. Les entités nommées sont réparties dans 4 classes:'. Below the text is a bulleted list of the four classes.

Entités nommées Istex

DESCRIPTION

Ce jeu représente la classification des entités nommées extraites des corpus ISTEX à l'aide du logiciel Unitex-CasSys. Les entités nommées sont réparties dans 4 classes:

- personnes
- lieux
- organisations
- indicateurs temporels

Quels fichiers choisir ?

Cas d'usage



2. Formats et types de fichiers

Créez votre sélection en cochant ou décochant les cases ci-dessous :

☐ Texte intégral 

☒ PDF

☐ TEI

☒ TXT

☐ ZIP

☐ TIFF



☐ Métadonnées 

☒ JSON

☐ XML

☐ MODS

☐ Annexes 

☐ Couvertures 

☐ Enrichissements 

☐ multicat

☐ nb

☐ refBibs

☐ teeft

☐ unitex

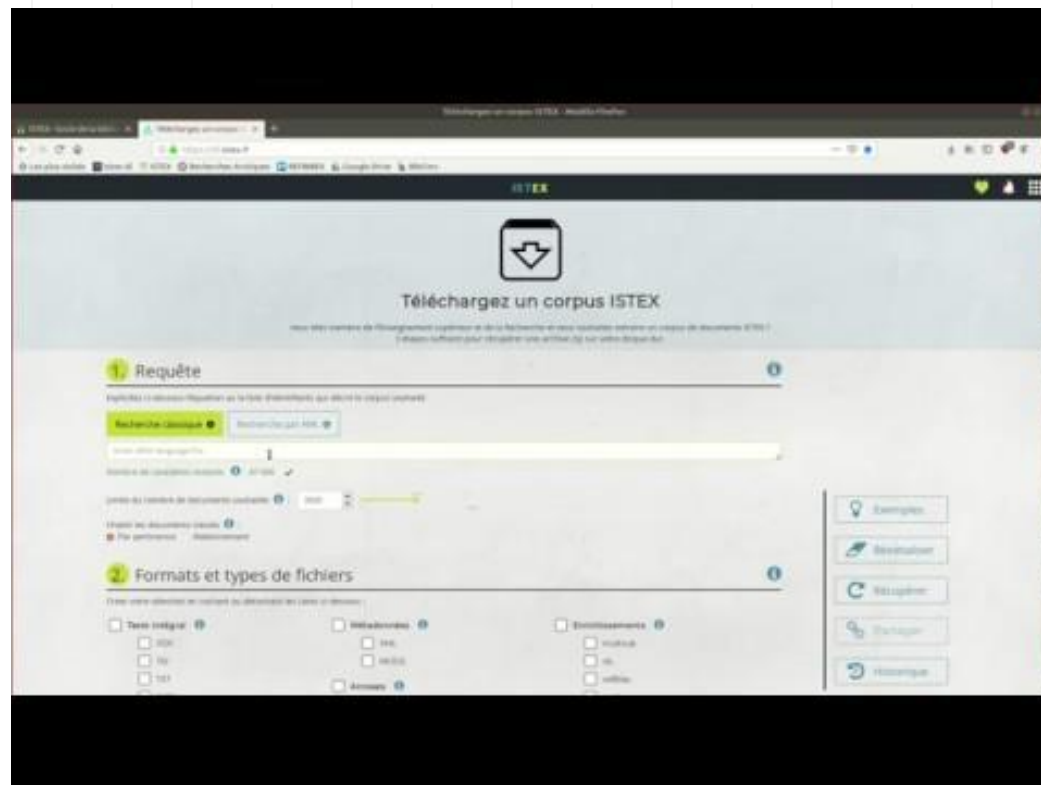
ISTEX-DL : étape 2

Fermeture
inopinée ?

Bouton
"Récupérer"



ISTEX-DL : étape 2



Fermeture
inopinée ?

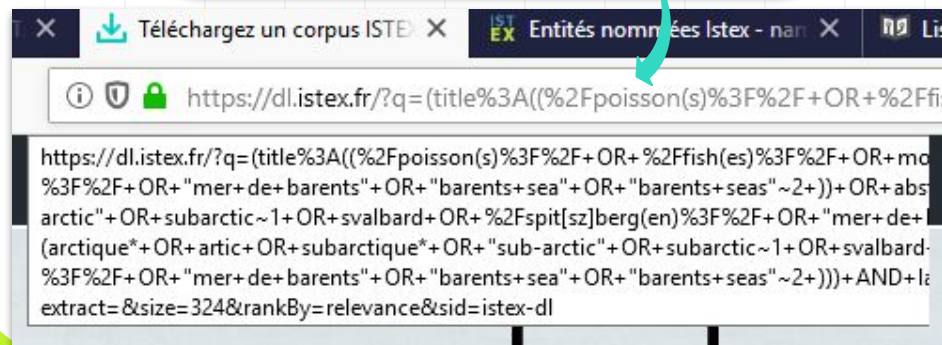
Bouton
"Récupérer"



À vérifier :

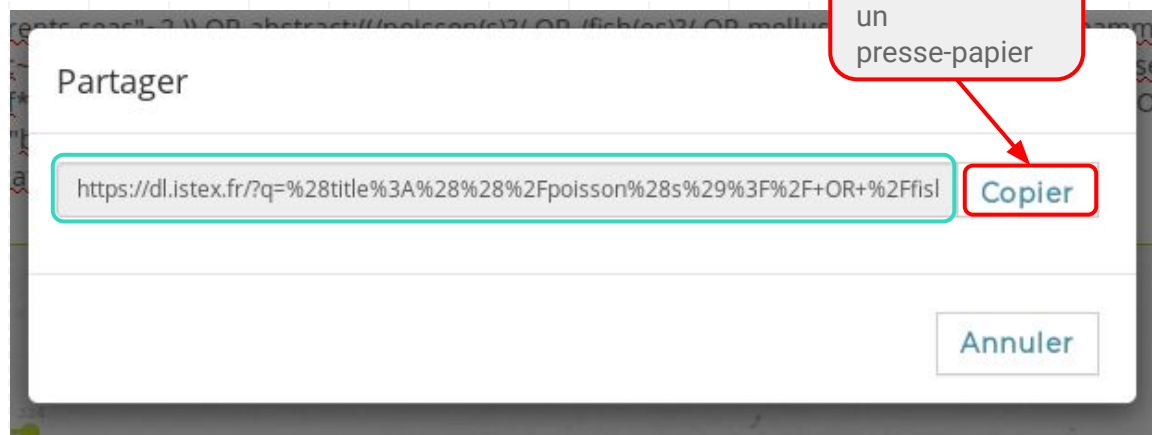
- Paramétrage navigateur : **autorisation** des cookies et des données de site jusqu'à leur expiration
- Connexion au site : **https://dl.istex.fr** uniquement, sans précision de paramètres (attention à l'adresse enregistrée en favori)

ISTEX-DL : étape 2



**Formulaire
complété...
2 boutons
s'activent**

**Bouton
"Partager"**



ISTEX-DL : étape 3

Formulaire
complété...
2 boutons
s'activent

Bouton
"Télécharger"



ISTEX-DL : étape 3

3. Télécharger

Lancer le
téléchargement...

Téléchargement en cours

La génération de votre corpus est en cours.
Veuillez patienter. L'archive sera bientôt téléchargée...



...et attendre la
réponse de l'API
ISTEX



Fermer

S'authentifier si
besoin

mode normal &
mode nomade

Authentification dans votre
établissement :

- automatique par
adresse IP connue

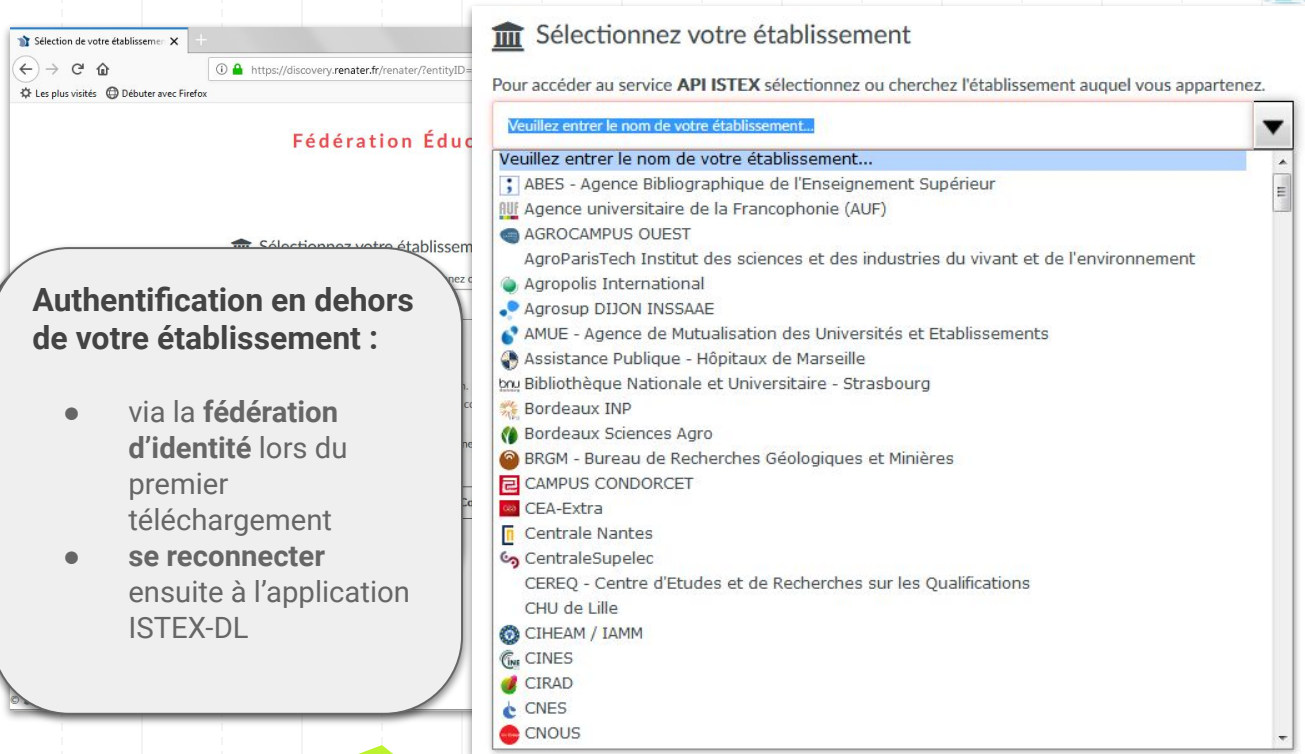
ISTEX-DL : étape 3

S'authentifier si
besoin

mode normal &
mode nomade

**Authentification en dehors
de votre établissement :**

- via la **fédération d'identité** lors du premier téléchargement
- **se reconnecter** ensuite à l'application ISTEK-DL



ISTEK-DL : étape 3

Fin du téléchargement

3. Télécharger



Enregistrer et attendre l'extraction complète

ISTEX-DL : étape 3

Ouverture de istex-subset-2019-

Vous avez choisi d'ouvrir :

 **istex-subset-2019-**

qui est un fichier de

à partir de : <https://>

Que doit faire Firefox av

☐ Ouvrir avec

Explor

☒ Enregistrer le fichier

☐ Toujours effectuer cette action

Téléchargement en cours

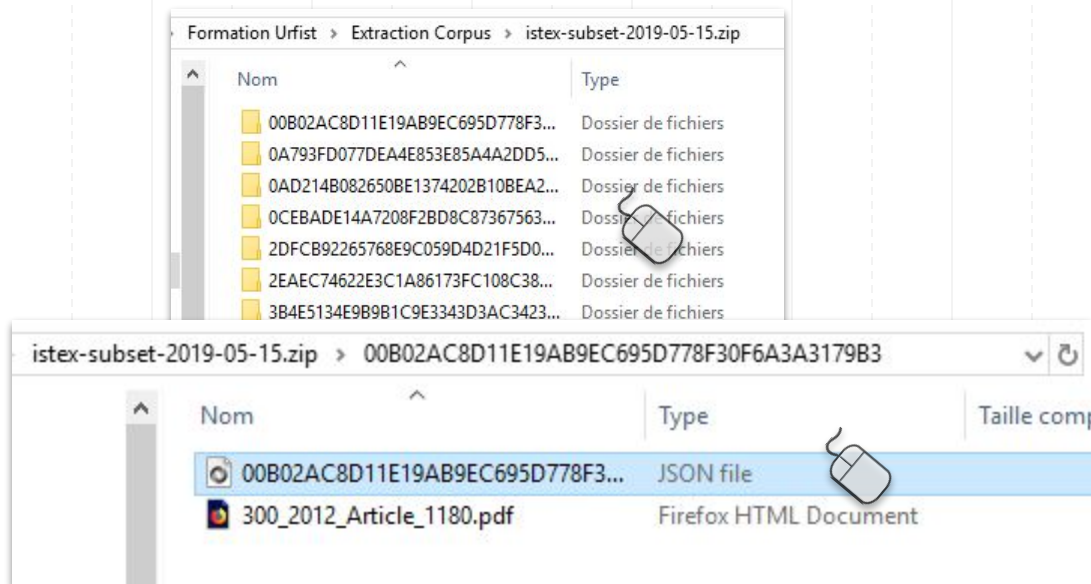
La génération de votre corpus est en cours.
Veuillez patienter. L'archive sera bientôt téléchargée...



Fermer

Ne pas oublier de fermer pour que le formulaire soit remis à zéro (les boutons partager et télécharger sont alors désactivés et l'historique mis à jour)

Fichier zip du corpus extrait



ISTEX-DL : étape 3

Fichier zip du corpus extrait

```
{
  "arkIstex": "ark:/67375/VQC-271B662S-H",
  "title": "Feeding ecology of dominant groundfish in the northern Bering Sea",
  "abstract": "Abstract: We investigated current diets of the six most abundant benthic fish in the northern Bering Sea. Our objective was to explore feeding strategies and potential competition with other top predators as ecosystem changes occur in the northern Bering Sea ecosystem. Our approach used stomach content data collected from field sampling during spring 2006 and 2007. Calanoid copepods and ampeliscid amphipods were important prey of Arctic cod (Boreogadus saida) but in different proportions depending upon fish size, feeding location, and local environmental conditions. Snailfish (Liparidae) occupied a broad niche and fed on a variety of benthic amphipods. Arctic alligatorfish (Ulcina olrikii) and Arctic staghorn sculpin (Gymnancistrus tricuspidis) consumed ampeliscid amphipods predominantly. Shorthorn sculpin (Myoxocephalus scorpius) had a less-diverse diet, with snow crab (Chionoecetes opilio) most important by weight. Finally, all Bering flounder (Hippoglossoides robustus) sampled had empty stomachs. Our results indicate that ampeliscid amphipods, which have high biomass in the central region of the northern Bering Sea, are the most important prey for the dominant groundfish in the Chirikov Basin. Generally, all dominant benthic fish in the northern Bering Sea had narrow feeding niches, except snailfish. High diet overlap was found among many of the fish species, including Arctic cod and snailfish, snailfish and Arctic alligatorfish, and Arctic alligatorfish and Arctic staghorn sculpin. These findings are consistent with a relatively short food chain for benthic fish that are for the most part specialized feeders with narrow preferences for food and may be affected by changes in benthic prey distributions.",
  "language": [
    "eng"
  ],
  "publicationDate": "2012",
  "copyrightDate": "2012",
  "doi": [
    "10.1007/s00300-012-1180-9"
  ],
  "ark": [
    "ark:/67375/VQC-271B662S-H"
  ],
  "articleId": [
    "1180",
    "s00300-012-1180-9"
  ],
  "genre": [
    "research-article"
  ],
  "originalGenre": [
    "OriginalPaper"
  ],
  "author": [
    {

```

ISTEX-DL : étape 3

Conserver une
trace
& Rejouer

Bouton
"Historique"



ISTEX-DL : étape 3

Historique des requêtes

#	Date	Requête	Formats	Nb. docs	Tri	Actions
1	Tue, 30 Apr 2019 11:52:14 GMT	(title:(/poisson(s)?/ OR /fish(es)?/ OR mollus* OR mammif* OR mammal*) AND (arctique* OR artic OR subarctique* OR "subarctic~1 OR svalbard OR /spit[sz]berg(en)?/ OR "mer de b		324	relevance	[Edit] [Download] [Share] [Close]
2	Tue, 30 Apr 2019 11:51:35 GMT	host.title.raw("Marine Biology") AND qualityIndicators.pdfVe TO *} AND qualityIndicators.score:[3.5 TO *} AND publication		232	random	[Edit] [Download] [Share] [Close]
3	Tue, 30 Apr 2019 11:50:19 GMT	title(sub?arctic + subarctic~1 + "norwegian seas"~2) OR abstr (sub?arctic + subarctic~1 + "norwegian seas"~2)		225	relevance	[Edit] [Download] [Share] [Close]
4	Tue, 30 Apr 2019 11:48:54 GMT	(species:genus:genera) AND ("plant genetic*" "plant taxonomy" "plant colonizant*" "phytogeograph*")) NOT ((fungi fungus funguses) NOT mycorrhiz*)	enrichme			[Edit] [Download] [Share] [Close]
5	Tue, 30 Apr 2019 11:41:40 GMT	(host:issn:"1822-6435" AND publicationDate:[1980 TO 2010]) OR (host.title:"Journal of Geophysical Research: Space Physics" AND host.issue.raw("1"))	metadata(xml,mods]			[Edit] [Download] [Share] [Close]

Supprimer l'historique

Fermer

4 actions possibles

Ré-éditer le formulaire pour le modifier

Télécharger le même corpus

Partager le même corpus

Effacer cette sauvegarde

Sauvegarde des dernières requêtes (30 max)

Aide



Documentation :

- Application ISTEEX-DL
 - adresse:
<https://doc.istex.fr/tdm/extraction/istex-dl.html>
 - directement via le pied de page

NB : Application en évolution constante...

Vos retours sont bienvenus !



Autres Outils : Quels Avantages ?

1 - Fonction "extract" (API ISTEX) :

à ajouter et à paramétrer directement dans l'URL d'interrogation

<https://api.istex.fr/document/?q=...&extract=...>

- option **rankBy = qualityOverRelevance**
- option **rankBy = random** avec **randomSeed** (nombre unique à 13 chiffres)

NB : ISTEX-DL exploite la fonction "extract"



Documentation

- Fonction "extract" :
<https://doc.istex.fr/tdm/extension/extract-feature.html>

Autres Outils : Quels Avantages ?

2 - Moissonneurs : utilitaires à installer et à lancer en ligne de commande avec la liste des options choisies :

"istex-api-harvester" :

- extraction **au-delà de 100 000** documents (scroll à paramétrer)
- tous modes d'option rankBy



Documentation

Tous les détails pour installer et utiliser :

- moissonneur **istex-api-harvester**

<https://doc.istex.fr/tdm/extension/istex-api-harvester.html>

Autres Outils : Quels Avantages ?

2 - Moissonneurs : utilitaires à installer et à lancer en ligne de commande avec la liste des options choisies :

“harvestCorpus” :

- extraction au-delà de 100 000 documents
- fichier **“.corpus”** (liste des identifiants des documents)
- **statsCorpus** : outil de calcul de **statistiques** associé à cet utilitaire
- tirage aléatoire (avec “graine aléatoire” possible)
- possibilité de renommer les fichiers extraits

```
#
# Fichier .corpus
#
title      : <à compléter>
author
publ
que
dat
li
ver
ran
tot
[IS
ar
ar
ar
ar
ark
ark a
...
```

Documentation

Tous les détails pour installer et utiliser :

- moissonneur **harvestCorpus**

<https://doc.istex.fr/tdm/extraction/harvest-corpus.pl.html>



Télécharger un corpus ISTEK > Et après ?

Exploration du corpus dans LODEX

Import du fichier dans Lodex

Modélisation à l'aide du modèle prédéfini

Exploration des différents graphiques

Objectifs :

- Explorer le contenu du corpus pour affiner à nouveau la requête
- Exposer le corpus pour le partager avec la communauté scientifique



Exploration et valorisation du corpus

LODEX

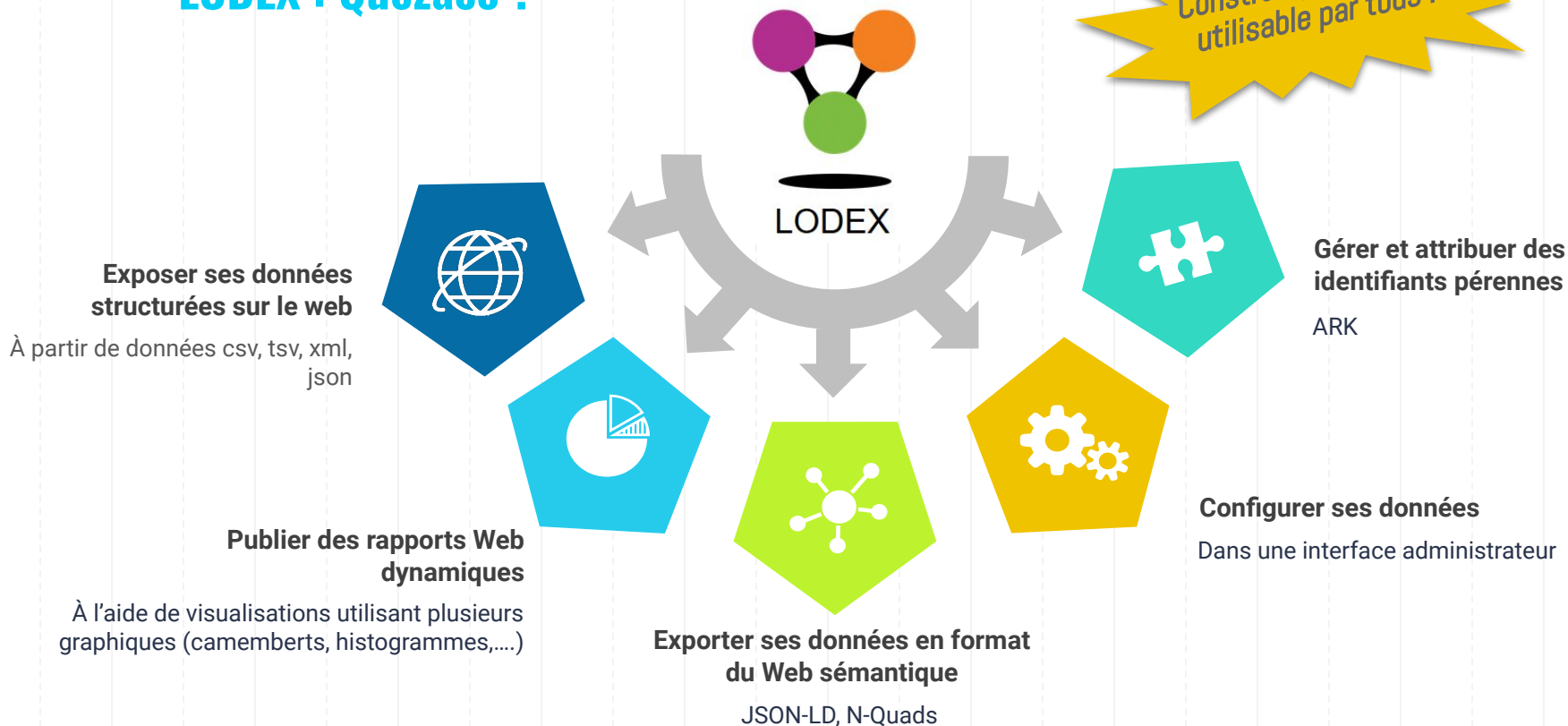
4



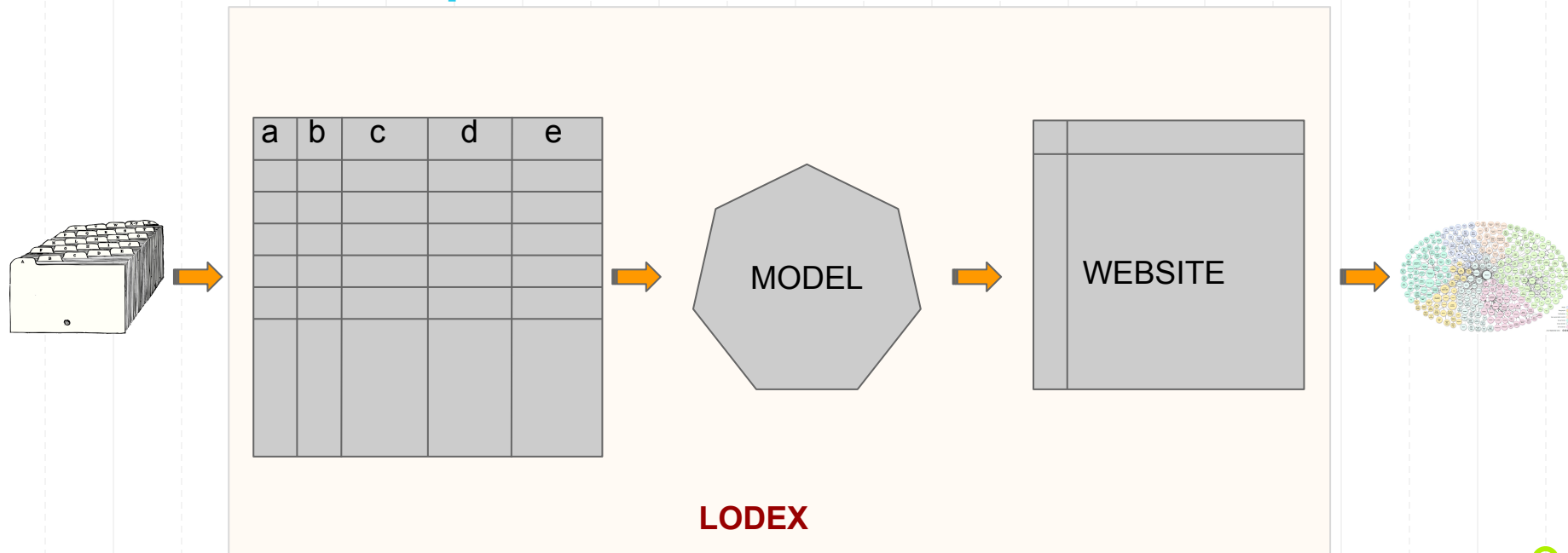
LODEX : Quézaco ?

Application web open source
dédiée aux données structurées

Construit pour ISTEEX,
utilisable par tous !



LODEX > Principe



LODEX > on passe à l'action !

1. Accéder à une instance de l'outil

- Adresse : <http://formation-urfistn.2.lodex-dev.inist.fr/>
- Nom d'utilisateur : **admin**
- Mot de passe : **secret**

instances numérotées
de 2 à 20



LODEX > on passe à l'action !

2. Import des données



COMMENCEZ EN CHARGEANT UN FICHIER DE DONNÉES.

1 Choix du parser à appliquer

Nom du parser

AUTO - avec l'extension du nom de fichier



Choisir dans le menu déroulant :
ZIP - résultat de
dl.istex.fr

2 Choix de fichier à charger

IMPORTER UN FICHIER

Choisir le fichier téléchargé
depuis ISTE-X-DL :
istex-subset-YYYY-MM-DD.zip

1 - Cliquer sur Voir Plus

2 - Cliquer sur Admin

Voir Plus

Admin

Déconnexion

LODEX > on passe à l'action !

Le jeu de données importé

Lodex - Modélisez et publiez vos données vers le web sémantique !

IMPORTER UN JEU DE DONNÉE

IMPORTER LE MODÈLE

EFFACER ▼

DÉCONNEXION

ERREURS

IMPORTÉ

Affiliation(s)	ARK	Auteur(s)	Auteur(s) monographie	Catégories INIST	Catégories Science-Metrix	Catégories Scopus	Catégories WoS	Chapitre
["Department of Natural S...	ark:/67375/WNG-CHKK1Z0N-H	["Jesper Hansen", "Nils-Mar...	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - ecology", "Cla...	[{"Nom": "3 - Ecology", "Cla...	[{"Nom": "2 - geography, ph...	
["Institute of Marine Res...	ark:/67375/VQC-74LKFOCS-R	["Elena Eriksen", "Bjarte B...	[]	[]	[]	[]	[]	
["Institut für Polarokolo...	ark:/67375/1BB-5CTZBF3P-Z	["C. F. von Dornien"]	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - marine biolog...	[{"Nom": "3 - General Agric...	[{"Nom": "2 - ecology", "Cla...	
["VNIRO, 17 V. Krasnosels...	ark:/67375/WNG-L6WPQ8B3-3	["V. M. Borisov", "A. A. El...	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - fisheries", "C...	[{"Nom": "3 - Aquatic Scien...	[{"Nom": "2 - marine & fres...	
[]	ark:/67375/HXZ-1MPJ41L6-4	["T.G. Sazykina"]	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - nuclear medic...	[{"Nom": "3 - Public Health...	[{"Nom": "2 - public, envir...	
["Carl von Ossietzky Univ...	ark:/67375/6H6-G3GH4DDW-3	["G.-P. Zauke", "V.M Savinov...	[]	[{"Nom": "4 - productions a...	[{"Nom": "3 - environmental...	[{"Nom": "3 - Pollution", "C...	[{"Nom": "2 - environmental...	

1197 ressou

56 colonnes chargées

nnes ajoutées

1 ressource = 1 document

1 colonne = 1 champ indexé

colonne ajoutée = URI obligatoire

LODEX > on passe à l'action !

1.1. Création du modèle

- Ajouter une colonne depuis le jeu de données original
- Ajouter une nouvelle colonne

1197 ressources chargées 56 colonnes chargées 62 colonnes ajoutées

uri	Nom du corpus	Nombre de documents	Titre de l'article	Lien vers le PDF	Type de corpus	Domaine(s)	Auteur(s)	Affiliation(s)	Langue
uid:/F8Z20TW8	Les poissons et les mollus...	/api/run/count-all	Late Pleistocene and Holoc...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["Jesper Hansen","Niels-Mar...	["Department of Natural S...	anglais
uid:/TZDP19FM	Les poissons et les mollus...	/api/run/count-all	Ecological significance of...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["Elena Eriksen","Bjarte B...	["Institute of Marine Res...	anglais
uid:/CQSKD1FG	Les poissons et les mollus...	/api/run/count-all	Reproduction and larval ec...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["C. F. von Dornien"]	["Institut für Polarökolo...	anglais
uid:/D297MCJP	Les poissons et les mollus...	/api/run/count-all	Long-term variations of bi...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["V. M. Borisov","A. A. El...	["VNIRO, 17 V. Krasnosels...	anglais
uid:/MD69C5G4	Les poissons et les mollus...	/api/run/count-all	Long-Distance Radionuclide...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["T.G. Sazykina"]	[]	anglais
uid:/T0T5PT3G	Les poissons et les mollus...	/api/run/count-all	Heavy metals in fish from ...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie","Géographie"]	["G.-P. Zauke","V.M Savinov...	["Carl von Ossietzky Univ...	anglais

LODEX > Le modèle

Permet de configurer toutes les informations à afficher sur un corpus, à partir :

- Du jeu de données
- De nouvelles colonnes

newField 2

- 1 Informations générales
- 2 Comment la valeur est créée
- 3 Transformations appliquées à cette valeur
- 4 Sémantiques
- 5 Comment et où elle est affichée
- 6 Recherche

LODEX > on passe à l'action !

1.2. Import d'un modèle existant



Importer un modèle existant remplacera le modèle actuel. Êtes-vous sûr de vouloir continuer ?

CONFIRMER

ANNULER

Confirmer et
Sélectionner le fichier
modele-atelier.json

Un même modèle peut être appliqué à différents jeux de données de structure identique

LODEX > on passe à l'action !

Le modèle importé

Lodex - Modélisez et publiez vos données vers le web sémantique !

IMPORTER UN JEU DE DONNÉE

IMPORTER LE MODÈLE

EFFACER

DÉCONNEXION

PUBLIER

IMPORTÉ

Affiliation(s)	ARK	Auteur(s)	Auteur(s) monographie	Catégories INIST	Catégories Science-Metrix	Catégories Scopus	Catégories WoS	Chapitre
["Department of Natural S...	ark:/67375/WNG-CHKK120N-H	["Jesper Hansen", "Nils-Mar...	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - ecology", "Cla...	[{"Nom": "3 - Ecology", "Cla...	[{"Nom": "2 - geography, ph...	
["Institute of Marine Res...	ark:/67375/VQC-74LKF0CS-R	["Elena Eriksen", "Bjarte B...	[]	[]	[]	[]	[]	
["Institut für Polarökolo...	ark:/67375/1BB-5CTZBF3P-Z	["C. F. von Dorrien"]	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - marine biolog...	[{"Nom": "3 - General Agric...	[{"Nom": "2 - ecology", "Cla...	
["VNIRO, 17 V. Krasnosels...	ark:/67375/WNG-L6WPQ8B3-3	["V. M. Borisov", "A. A. El...	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - fisheries", "C...	[{"Nom": "3 - Aquatic Scien...	[{"Nom": "2 - marine & fres...	
[]	ark:/67375/HXZ-1MPJ41L6-4	["T.G. Sazykina"]	[]	[{"Nom": "3 - sciences biol...	[{"Nom": "3 - nuclear medic...	[{"Nom": "3 - Public Health...	[{"Nom": "2 - public, envir...	
["Carl von Ossietzky Univ...	ark:/67375/6H6-G3GH4DDW-3	["G.-P. Zauke", "V.M Savinov...	[]	[{"Nom": "4 - productions a...	[{"Nom": "3 - environmental...	[{"Nom": "3 - Pollution", "C...	[{"Nom": "2 - environmental...	

1197 ressources chargées

56 colonnes chargées

62 colonnes ajoutées

+

APERÇU

uri	Nom du corpus	Nombre de documents	Titre de l'article	Lien vers le PDF	Type de corpus	Domaine(s)	Auteur(s)	Affiliation(s)	Langue(s)
uid:/F8220TW8	Les poissons et les mollus...	/api/run/count-all	Late Pleistocene and Holoc...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie", "Géographie"]	["Jesper Hansen", "Nils-Mar...	["Department of Natural S...	anglais
uid:/TZDP19FM	Les poissons et les mollus...	/api/run/count-all	Ecological significance of...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie", "Géographie"]	["Elena Eriksen", "Bjarte B...	["Institute of Marine Res...	anglais
uid:/CQSKD1FG	Les poissons et les mollus...	/api/run/count-all	Reproduction and larval ec...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie", "Géographie"]	["C. F. von Dorrien"]	["Institut für Polarökolo...	anglais
uid:/D297MCJP	Les poissons et les mollus...	/api/run/count-all	Long-term variations of bi...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie", "Géographie"]	["V. M. Borisov", "A. A. El...	["VNIRO, 17 V. Krasnosels...	anglais
uid:/MD69C5G4	Les poissons et les mollus...	/api/run/count-all	Long-Distance Radionuclide...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie", "Géographie"]	["T.G. Sazykina"]	[]	anglais
uid:/T0T5PT3G	Les poissons et les mollus...	/api/run/count-all	Heavy metals in fish from ...	https://api.istex.fr/ark:/...	Corpus thématique	["Zoologie", "Géographie"]	["G.-P. Zauke", "V.M Savinov...	["Carl von Ossietzky Univ...	anglais

LODEX > on passe à l'action !

Les données publiées

Lodex - Modélisez et publiez vos données vers le web sémantique !

AJOUTER DES DONNÉES

MODÉRATION

RESSOURCES RETIRÉES

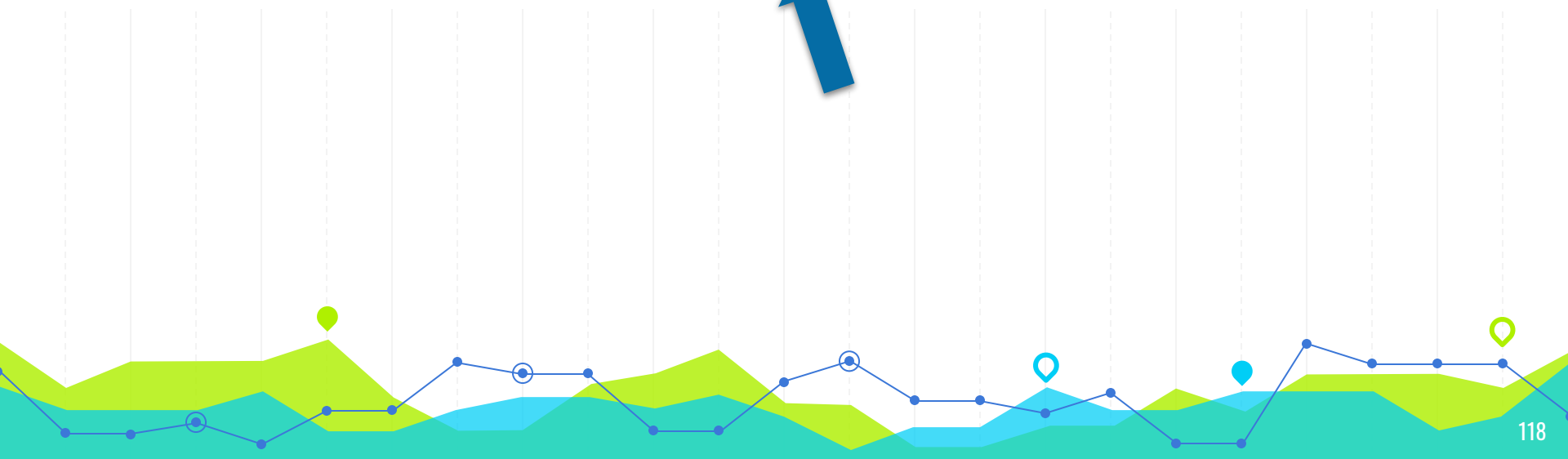
VOIR LE MODÈLE

EFFACER ▼

DÉCONNEXION

Vos données ont été publiées en ligne. Vous pouvez configurer davantage l'affichage de ces dernières ou les mettre à jour en utilisant l'interface publique

ALLER AUX DONNÉES PUBLIÉES



LODEX > comment naviguer dans le corpus

Accès à la page d'accueil



Accès aux graphiques
avec filtre par facettes



Recherche de documents
avec filtre par facettes



Accès à d'autres
fonctionnalités



LODEX > exploration des données publiées



Page d'accueil

- Présentation du corpus

- Identification du corpus
- Rappel du cas d'usage
- Possibilité de rejouer la requête
- Droits d'utilisation des documents
- Citation du corpus



Exposition du corpus

LODEX > modification des données publiées

- 2 icônes disponibles pour chaque champ



Pour modifier la valeur du champ



Pour modifier le paramétrage du champ (format, facette, etc.)

A yellow lightbulb icon with rays emanating from it, indicating a tip or important note.

*Permet d'adapter le modèle
d'un corpus à l'autre*

- Exercices

1. Modifier le titre du corpus
2. Modifier le format du titre du corpus

LODEX > exploration des données publiées



Graphiques

Les graphiques

- Visualisation de la répartition des documents selon différents angles de vue
 - Dates de publication
 - Revues
 - Type de publication, type de document
 - Éditeurs
 - Mots-clés d'auteur
 - Catégories scientifiques (Science-Metrix, Scopus, WoS, Inist)
 - Indicateurs (version de PDF, score de qualité)



Analyse du corpus

LODEX > exploration des données publiées

Les graphiques

• Exercices

1. Explorer le graphique "Revues"

- ☐ utiliser la facette "Titre de revue" pour accéder aux documents d'une revue
- ☐ utiliser la facette "Date de publication" pour cibler les revues récentes
- ☐ changer le nombre de revues affichées

LODEX > exploration des données publiées

Les graphiques

- Exercices

2. Explorer le graphique "Type de documents"

☐ utiliser la facette "Type de documents" pour identifier les documents qui ne sont **pas** de type : article, article de recherche, article de synthèse

3. Explorer le graphique "Catégories Science-Metrix"

☐ naviguer dans les différents niveaux de hiérarchie pour valider la qualité du corpus

LODEX > de l'exploration à la ré-interrogation d'ISTEX

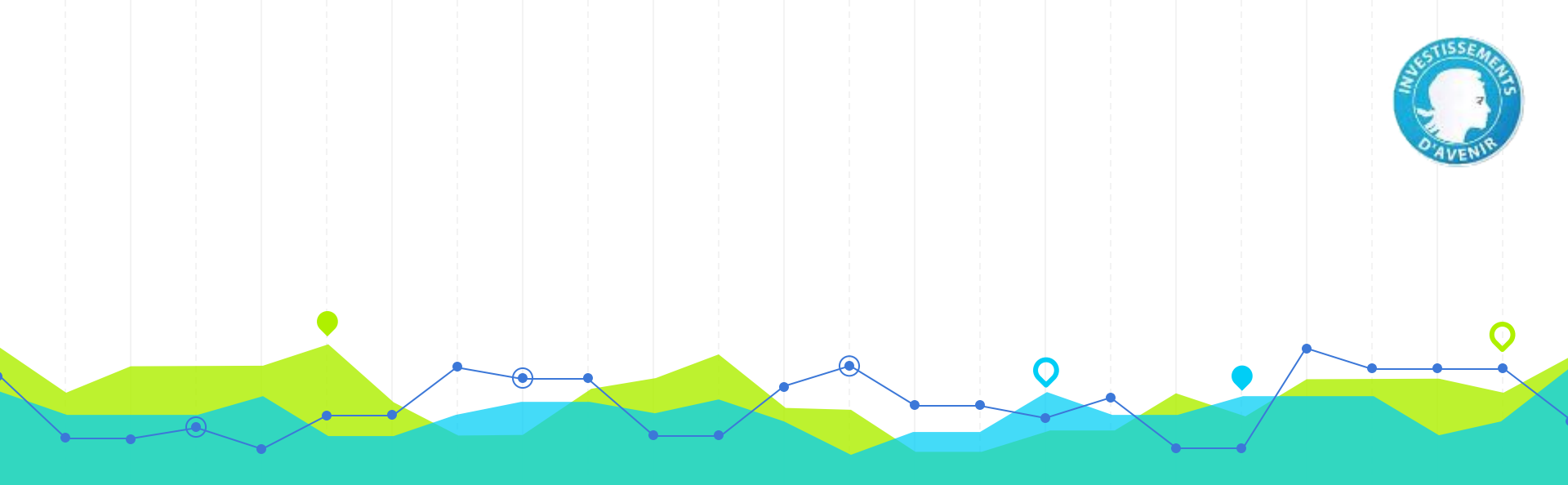
Équation affinée

```
(title:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))  
  
OR abstract:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 ))  
  
OR subject.value:(/fish(es)?/ OR mollus*) AND (arctic OR "sub-arctic" OR subarctic~1 OR svalbard OR  
/spit[sz]berg(en)?/ OR "barents sea" OR "barents seas"~2 )))  
  
AND language:eng  
  
AND qualityIndicators.pdfWordCount:[500 TO 10000]  
  
NOT genre:abstract NOT categories.scienceMetrix:pediatrics
```

Résultats (03 mars 2020): 1 197 ➤ 1 194 docs

LODEX > exemples de corpus Inist

- Corpus avec enrichissement classification issue de Topic Modeling
 - [Vieillissement](#)
- Corpus avec annotation manuelle disponible
 - [Unitex](#)
- Corpus avec enrichissement systématique animale
 - [Échantillon Systématique Animale](#) (version en développement)
- Corpus avec évolution diachronique des thématiques
 - [Géosciences](#) (version en développement)



Liens utiles

Adresses & Co
Documentation & Informations
Contact & Discussion

5

Adresses & Co



Se connecter :

- ISTEK : <http://www.istex.fr>
- Démonstrateur ISTEK : <http://demo.istex.fr/>
- Interface ISTEK-DL : <https://dl.istex.fr/>
- Infos Lodex : <https://lodex.inist.fr/>
- Données ISTEK : <https://data.istex.fr/>

S'authentifier :

- Vérifier ses droits d'accès : <https://api.istex.fr/auth>
- Vérifier son accès par fédération d'identité :
<https://api.istex.fr/auth?auth=fede>

Documentation & Informations



Se documenter :

- Documentation "Usage TDM d'ISTEX" : <https://doc.istex.fr/tdm/>
- Documentation "API ISTEX" : <https://doc.istex.fr/api/>
- Documentation "Lodex" : <https://user-doc.lodex.inist.fr/>

Se tenir informé :

- Blog ISTEX : <https://blog.istex.fr/>
- Plateforme Twitter : [@ISTEX_Platform](#)

Contact & Discussion



Chercher de l'aide / Contribuer à l'amélioration :

- Contact :
 - Via le formulaire : <https://www.istex.fr/contact/>
 - Via la liste : contact@listes.istex.fr
- Liste de discussion : users@listes.istex.fr (liste publique)

L'équipe ISTEX





Annexes 6

S'aider...



Interrogation > quelques champs utiles



Penser à utiliser les champs **"host.title"** et **"host.issn"** pour rechercher une revue (changement d'ISSN ou de titre au cours du temps, erreurs orthographiques, etc.)

Titre d'article	title
Auteur	author.name
Résumé	abstract
Mots-clés d'auteur	subject.value
Date de publication	publicationDate
Langue	language
Titre de revue	host.title
ISSN revue	host.issn



Les champs **.raw** permettent d'interroger sur la valeur exacte du champ.

Exemple :
q=host.title.raw:"Science"
ne ramènera que les documents de la revue Science et non ceux de "The Science of the Total Environment", etc.

S'aider...

Interrogation > quelques champs utiles

	Éditeur	corpusName
	Type de publication	host.genre
	Type de document	genre
Enrichissements		
Disciplines scientifiques		categories.scopus, categories.inist, categories.sciencesMetrix, categories.wos
Entités nommées		namedEntities.unitex, etc.

S'aider...



Interrogation > quelques champs utiles

Indicateurs (TDM)

Score de qualité

`qualityIndicators.score`

Version de PDF

`qualityIndicators.pdfVersion`

Nombre de mots par PDF

`qualityIndicators.pdfWordCount`

À venir

Pdf image

`qualityIndicators.pdfText`

Nombre de mots par page

`qualityIndicators.pdfWordsPerPage`

Sauter
le pas...



Interrogation > intérêts de l'API ISTEX

- Exploration du contenu du fonds ISTEX et répartition des documents en fonction d'une facette
Exemple : connaître toutes les langues des documents présents dans ISTEX
- Interrogation selon une ou plusieurs facettes imbriquées
Exemple : Pour chaque langue présente dans ISTEX, connaître le nombre de documents par siècle
- Interrogation de toutes les facettes disponibles dans l'API (nombre réduit sur le démonstrateur)



**Interface de Programmation
Applicative** : solution informatique
qui permet à des applications de
communiquer entre elles et de
s'échanger mutuellement des
services ou des données

Tutos

Pour comprendre ce qu'est l'API
ISTEX et apprendre à l'utiliser,
reportez-vous aux tutoriels :

<https://istex-tutorial.data.istex.fr/>

S'inspirer...



Interrogation > Exemples de requêtes sur l'API

*Exemple 1 : lister tous les éditeurs présents dans ISTEK et connaître le nombre de documents pour chacun d'eux **sans afficher** les documents (emploi de **size=0**)*

Éditeurs

[https://api.istex.fr/document/?q=*&facet=corpusName\[*\]&size=0](https://api.istex.fr/document/?q=*&facet=corpusName[*]&size=0)

Types de publication

[https://api.istex.fr/document/?q=*&facet=host.genre\[*\]&size=0](https://api.istex.fr/document/?q=*&facet=host.genre[*]&size=0)

Types de document

[https://api.istex.fr/document/?q=*&facet=genre\[*\]&size=0](https://api.istex.fr/document/?q=*&facet=genre[*]&size=0)

S'inspirer...



Interrogation > Exemples de requêtes sur l'API

Exemple 1 : explorer la répartition du nombre de documents Istex par période de 100 ans entre les 15^e et 21^e siècles



Dates de publication

par siècle (entre 1401 et 2100)

[https://api.istex.fr/document/?q=*&size=0&facet=publicationDate\[1401-2100:100\]](https://api.istex.fr/document/?q=*&size=0&facet=publicationDate[1401-2100:100])

par période de 10 ans (20^e et début 21^e siècles)

[https://api.istex.fr/document/?q=*&size=0&facet=publicationDate\[1901-2020:10\]](https://api.istex.fr/document/?q=*&size=0&facet=publicationDate[1901-2020:10])

Langues & dates

les dates pour chaque langue (facettes imbriquées)

[https://api.istex.fr/document/?q=*&facet=language\[*\]>publicationDate\[1401-2100:100\]&size=0](https://api.istex.fr/document/?q=*&facet=language[*]>publicationDate[1401-2100:100]&size=0)

S'inspirer...



Interrogation > Exemples de requêtes sur l'API

Exemples 1 & 2 : savoir si les documents présents dans ISTEEX **comportent ou non un résumé** et afficher ces documents



Présence de
résumé

avec résumé

[https://api.istex.fr/document/?q=\(abstract:*\)](https://api.istex.fr/document/?q=(abstract:*))

sans résumé

[https://api.istex.fr/document/?q=\(NOT \(abstract:*\)\)](https://api.istex.fr/document/?q=(NOT (abstract:*)))

Présence de
mots-clés d'auteur

avec mots-clés

[https://api.istex.fr/document/?q=\(subject.value.raw:*\)&size=0](https://api.istex.fr/document/?q=(subject.value.raw:*)&size=0)

sans mots-clés

[https://api.istex.fr/document/?q=\(NOT \(subject.value.raw:*\)&size=0](https://api.istex.fr/document/?q=(NOT (subject.value.raw:*)&size=0)

S'inspirer...



Interrogation > Exemples de requêtes sur l'API

Exemple 1 : quantifier les documents appartenant à **chaque catégorie Scopus** d'un corpus de documents sur l'Arctique



Domaines
scientifiques

catégories
Scopus d'un
corpus particulier

[https://api.istex.fr/document/?q=arctic&facet=categories.scopus\[*\]&size=0](https://api.istex.fr/document/?q=arctic&facet=categories.scopus[*]&size=0)

catégories
Science-Metrix de
l'archive ISTE

[https://api.istex.fr/document/?q=* &facet=categories.scienceMetrix\[*\]&size=0](https://api.istex.fr/document/?q=* &facet=categories.scienceMetrix[*]&size=0)

S'inspirer...

Interrogation > Exemples de requêtes sur l'API



Exemple 1 : obtenir la distribution du nombre de documents ISTEK en fonction de leur **score de qualité**, et ce par **pas de 1**

Indicateurs

score de qualité

[https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.score\[1-10:1\]](https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.score[1-10:1])

version de PDF

[https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.pdfVersion\[*\]](https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.pdfVersion[*])

nombre de mots
par PDF

[https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.pdfWordCount\[1-10000:1000\]](https://api.istex.fr/document/?q=*&size=0&facet=qualityIndicators.pdfWordCount[1-10000:1000])

Merci !

Avez-vous des questions ?

