

La constitution et l'exploration d'un corpus documentaire à partir des ressources Istex

Mathilde Huguin

IR PhD Inist - CNRS

mathilde.huguin@inist.fr

ANF TDM 2024 | Exploration documentaire et extraction d'informations



Inist- CNRS

Plan

1. Présentation d'Istex

Un peu d'hISToire

Objectifs & accès

Quelques chiffres

Chaîne de traitements

Présentation des outils

Istex dans le paysage des réservoirs

2. Constitution d'un corpus spécialisé

Cas d'usage

Méthodologie

Requêtage et téléchargement

Exploration et visualisation

3. Conclusion

4. Liens, contacts, références



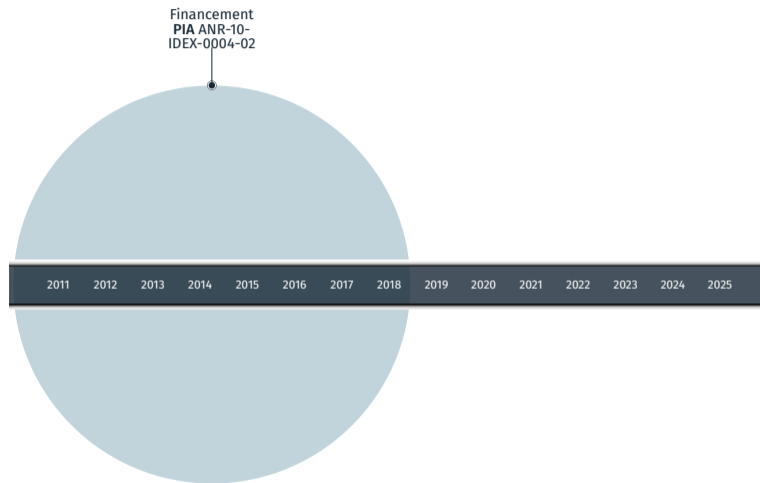
1. Présentation d'Istex

« Initiative d'excellence en Information Scientifique
et Technique »



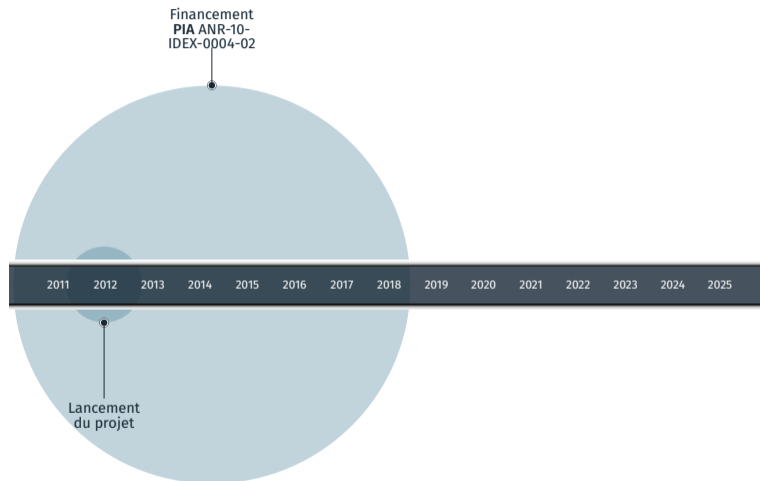
Présentation d'Istex

UN PEU D'HISTOIRE



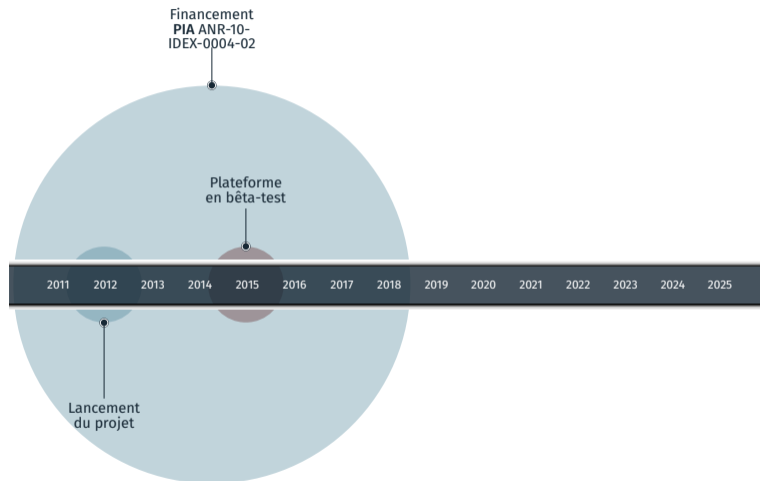
Présentation d'Istex

UN PEU D'HISTOIRE



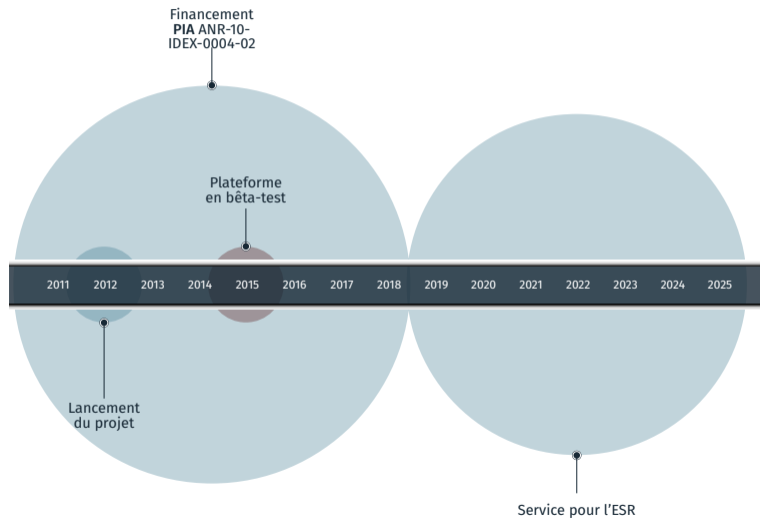
Présentation d'Istex

UN PEU D'HISTOIRE



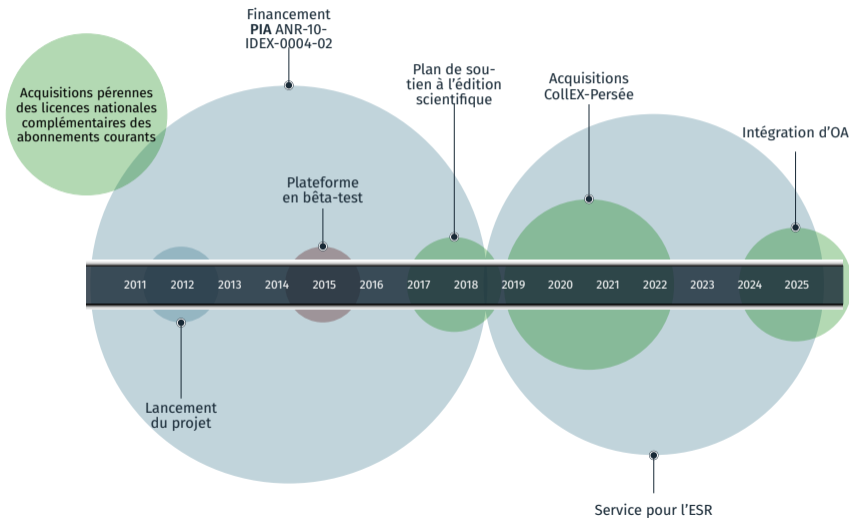
Présentation d'Istex

UN PEU D'HISTOIRE



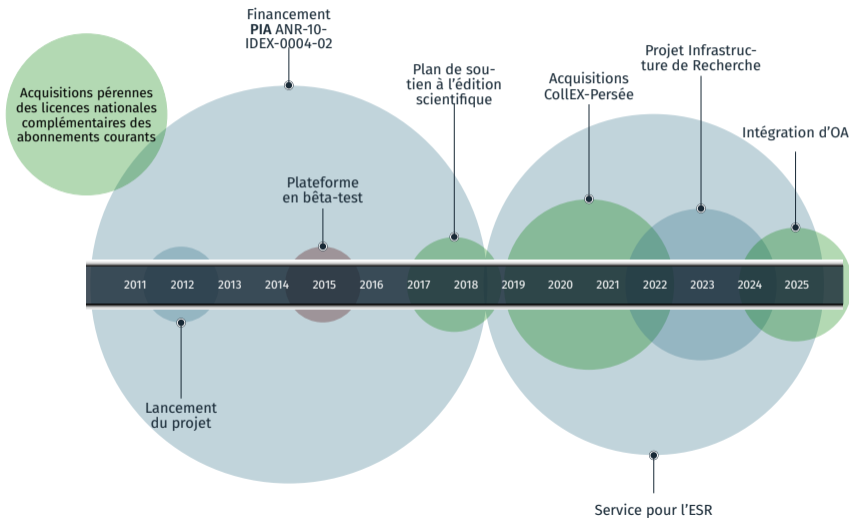
Présentation d'Istex

UN PEU D'HISTOIRE



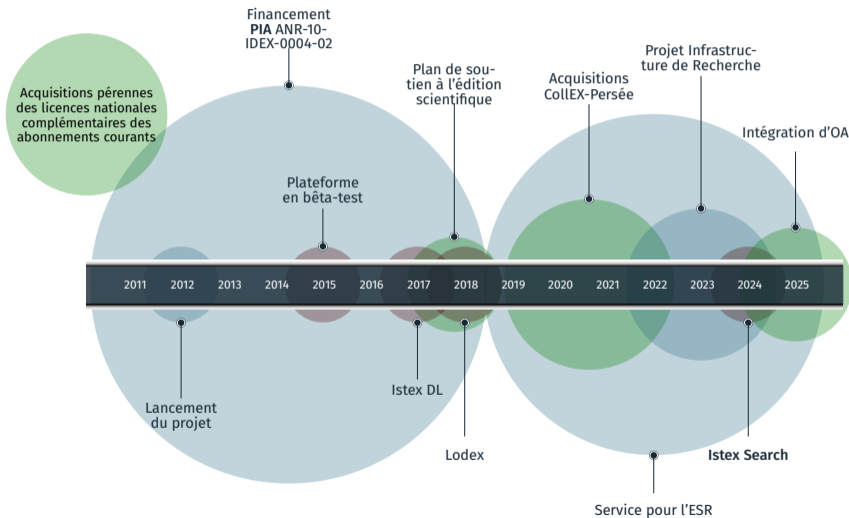
Présentation d'Istex

UN PEU D'HISTOIRE



Présentation d'Istex

UN PEU D'HISTOIRE



Présentation d'Istex

OBJECTIFS & ACCÈS



- Acquérir massivement des archives scientifiques multilingues et multidisciplinaires (cf. Bérard, 2012 ; Petitjean-Monnin *et al.*, 2022)

Présentation d'Istex

OBJECTIFS & ACCÈS



- Acquérir massivement des archives scientifiques multilingues et multidisciplinaires (cf. Bérard, 2012 ; Petitjean-Monnin *et al.*, 2022)
- Enrichir les données

Présentation d'Istex

OBJECTIFS & ACCÈS



- Acquérir massivement des archives scientifiques multilingues et multidisciplinaires (cf. Bérard, 2012 ; Petitjean-Monnin *et al.*, 2022)
- Enrichir les données
- Mettre à disposition les données via une plateforme nationale (cf. Niederlender, 2017)

Présentation d'Istex

OBJECTIFS & ACCÈS



- Acquérir massivement des archives scientifiques multilingues et multidisciplinaires (cf. Bérard, 2012 ; Petitjean-Monnin *et al.*, 2022)
- Enrichir les données
- Mettre à disposition les données via une plateforme nationale (cf. Niederlender, 2017)



- Métadonnées accessibles à toutes et tous

Présentation d'Istex

OBJECTIFS & ACCÈS



- Acquérir massivement des archives scientifiques multilingues et multidisciplinaires (cf. Bérard, 2012 ; Petitjean-Monnin *et al.*, 2022)
- Enrichir les données
- Mettre à disposition les données via une plateforme nationale (cf. Niederlender, 2017)



- Métadonnées accessibles à toutes et tous
- Textes intégraux et enrichissements réservés à l'ESR (380 établissements adhérents)

Présentation d'Istex

OBJECTIFS & ACCÈS



- Acquérir massivement des archives scientifiques multilingues et multidisciplinaires (cf. Bérard, 2012 ; Petitjean-Monnin *et al.*, 2022)
- Enrichir les données
- Mettre à disposition les données via une plateforme nationale (cf. Niederlender, 2017)



- Métadonnées accessibles à toutes et tous
- Textes intégraux et enrichissements réservés à l'ESR (380 établissements adhérents)
- Lien avec des ENT, Google Scholar, [Click & Read](#)

Quelques chiffres*



*En date du 27 août 2024

Quelques chiffres*

27 950 833 documents

*En date du 27 août 2024

Quelques chiffres*

27 950 833 documents

49 bouquets d'éditeurs

*En date du 27 août 2024

Quelques chiffres*

27 950 833 documents

49 bouquets d'éditeurs

9 492 revues

*En date du 27 août 2024

Quelques chiffres*

27 950 833 documents

49 bouquets d'éditeurs

9 492 revues

438 666 monographies

*En date du 27 août 2024

Quelques chiffres*

27 950 833 documents

49 bouquets d'éditeurs

9 492 revues

438 666 monographies

700 ans de publications

*En date du 27 août 2024

Quelques chiffres*

27 950 833 documents

49 bouquets d'éditeurs

9 492 revues

438 666 monographies

700 ans de publications

60 langues

*En date du 27 août 2024

Présentation d'Istex

DEUX USAGES

Une ressource documentaire

- Trouver des articles scientifiques

Présentation d'Istex

DEUX USAGES

Une ressource documentaire

- Trouver des articles scientifiques
- Alimenter des bases documentaires

Présentation d'Istex

DEUX USAGES

Une ressource documentaire

- Trouver des articles scientifiques
- Alimenter des bases documentaires

Un matériau de recherche

- Constituer des corpus

Présentation d'Istex

DEUX USAGES

Une ressource documentaire

- Trouver des articles scientifiques
- Alimenter des bases documentaires

Un matériau de recherche

- Constituer des corpus
- Faire du TDM, de la scientométrie, du TAL (e.g. Atanassova et Bertin, 2023 ; Bannour *et al.*, 2024 ; Rauf et Yvon, 2024)

Présentation d'Istex

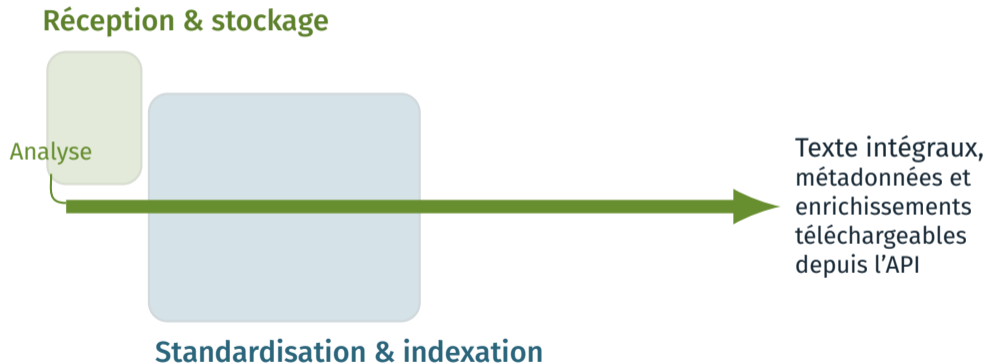
CHAÎNE DE TRAITEMENTS

Réception & stockage



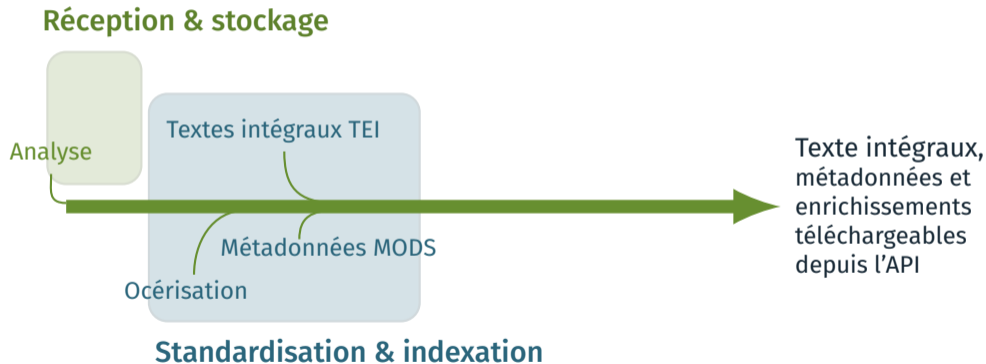
Présentation d'Istex

CHAÎNE DE TRAITEMENTS



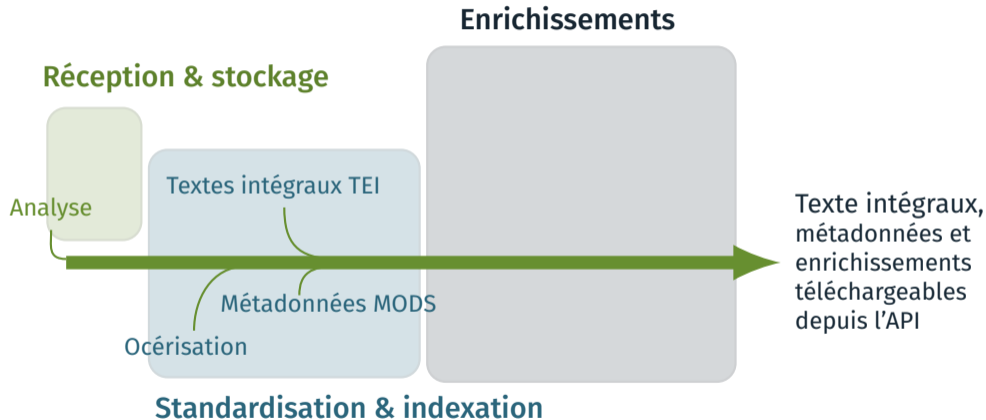
Présentation d'Istex

CHAÎNE DE TRAITEMENTS



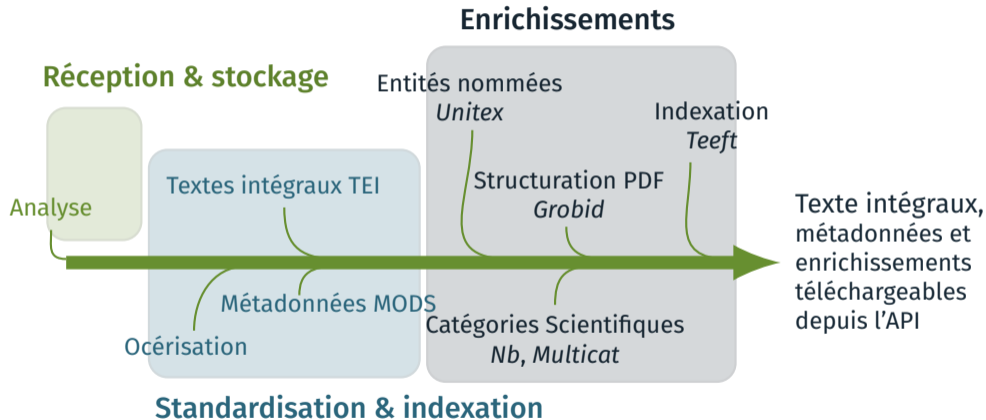
Présentation d'Istex

CHAÎNE DE TRAITEMENTS



Présentation d'Istex

CHAÎNE DE TRAITEMENTS



Présentation d'Istex

PRÉSENTATION DES OUTILS

ISTEX

Istex

Site général, vitrine du possible

www.istex.fr

Présentation d'Istex

PRÉSENTATION DES OUTILS

ISTEX

Istex

Site général, vitrine du possible

www.istex.fr



Data

Informations sur les données Istex (collections, enrichissements) & accès aux corpus (Cuxac *et al.*, 2019)

data.istex.fr

Présentation d'Istex

PRÉSENTATION DES OUTILS

ISTEX

Istex

Site général, vitrine du possible

www.istex.fr



Data

Informations sur les données Istex (collections, enrichissements) & accès aux corpus (Cuxac *et al.*, 2019)

data.istex.fr



Search

Application pour interroger l'API et télécharger des corpus

search.istex.fr

Présentation d'Istex

PRÉSENTATION DES OUTILS

ISTEX

Istex

Site général, vitrine du possible

www.istex.fr



Data

Informations sur les données Istex (collections, enrichissements) & accès aux corpus (Cuxac *et al.*, 2019)

data.istex.fr



Search

Application pour interroger l'API et télécharger des corpus

search.istex.fr







Lodex

Outil de création de sites web (visualisation / enrichissement de données) (Gregorio *et al.*, 2019)

www.lodex.fr







Présentation d'Istex

PRÉSENTATION DES OUTILS

ISTEX	Istex	Site général, vitrine du possible	www.istex.fr
	Data	Informations sur les données Istex (collections, enrichissements) & accès aux corpus (Cuxac <i>et al.</i> , 2019)	data.istex.fr
	Search	Application pour interroger l'API et télécharger des corpus	search.istex.fr
	Lodex	Outil de création de sites web (visualisation / enrichissement de données) (Gregorio <i>et al.</i> , 2019)	www.lodex.fr
	TDM	Bibliothèque de web services (Cuxac, 2024)	services.istex.fr

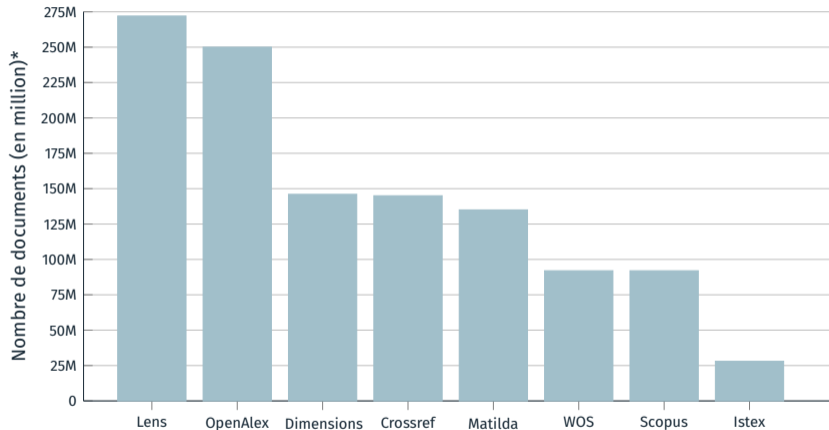
Présentation d'Istex

PRÉSENTATION DES OUTILS

	Istex	Site général, vitrine du possible	www.istex.fr
	Data	Informations sur les données Istex (collections, enrichissements) & accès aux corpus (Cuxac <i>et al.</i> , 2019)	data.istex.fr
	Search	Application pour interroger l'API et télécharger des corpus	search.istex.fr
	Lodex	Outil de création de sites web (visualisation / enrichissement de données) (Gregorio <i>et al.</i> , 2019)	www.lodex.fr
	TDM	Bibliothèque de web services (Cuxac, 2024)	services.istex.fr
	Loterre	Bibliothèque de terminologies multilingues (Khayari <i>et al.</i> , 2021)	www.loterre.fr

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | QUANTITÉ



Comparaison quantitative des réservoirs

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

Réservoirs	Origine des données	Fournisseurs
Lens	Microsoft Academic, PubMed, Crossref, OpenAlex, Unpaywall, ORCID	Cambia
OpenAlex	Microsoft Academic, Crossref, PubMed, Unpaywall, DOAJ, ORCID, ROR, Internet archive, crawls web, arXiv, Zenodo	OurResearch
Dimensions Matilda	Crossref, PubMed, DataCite, arXiv + 130 éditeurs Crossref, PubMed Central, arXiv et RePEc, BASE + Unpaywall, ORCID	Digital Science I3 & Huma-Num
Crossref	Crossref	Crossref
WOS	Revue sélectionnées	Clarivate
Scopus	Revue sélectionnées	Elsevier
Istex	Négociations, Achats, Plan de soutien, GIS CollEX-Persée, OA	Istex

Origine des données (Bouchard, 2024, Gusenbauer, 2022)

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+
Indicateurs bibliométriques	+	+	+	+	-

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+
Indicateurs bibliométriques	+	+	+	+	-
Indicateurs TDM	-	-	-	-	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+
Indicateurs bibliométriques	+	+	+	+	-
Indicateurs TDM	-	-	-	-	+
Texte intégral téléchargeable	-	-	-	-	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+
Indicateurs bibliométriques	+	+	+	+	-
Indicateurs TDM	-	-	-	-	+
Texte intégral téléchargeable	-	-	-	-	+
Formats TDM standardisés (e.g. TEI)	-	-	-	-	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+
Indicateurs bibliométriques	+	+	+	+	-
Indicateurs TDM	-	-	-	-	+
Texte intégral téléchargeable	-	-	-	-	+
Formats TDM standardisés (e.g. TEI)	-	-	-	-	+
Passerelles	-	-	-	-	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | PROPRIÉTÉS

	Lens	OpenAlex	Dimensions	Matilda	Istex
Filtres	+	+	+	+/-	+
Tris	+	+	+	+	+
Enrichissements	+	+	+	+	++
Recherche assistée	++	+/-	-	-	++
Formats bibliographiques (e.g. BIB, MODS)	+	+	+	+	+
Indicateurs bibliométriques	+	+	+	+	-
Indicateurs TDM	-	-	-	-	+
Texte intégral téléchargeable	-	-	-	-	+
Formats TDM standardisés (e.g. TEI)	-	-	-	-	+
Passerelles	-	-	-	-	+
Archives	-	-	-	-	+

Propriétés des réservoirs : des usages différents

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources
- ⊕ Des données interopérables, des formats homogénéisés et des données corrigées (moins de prétraitements)

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources
- ⊕ Des données interopérables, des formats homogénéisés et des données corrigées (moins de prétraitements)
- ⊕ Des données enrichies (réocérisation, structuration de texte, métadonnées) : des documents retrouvés et analysés plus facilement

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources
- ⊕ Des données interopérables, des formats homogénéisés et des données corrigées (moins de prétraitements)
- ⊕ Des données enrichies (réocérisation, structuration de texte, métadonnées) : des documents retrouvés et analysés plus facilement
- ⊕ Des millions de textes et de métadonnées téléchargeables facilement

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources
- ⊕ Des données interopérables, des formats homogénéisés et des données corrigées (moins de prétraitements)
- ⊕ Des données enrichies (réocérisation, structuration de texte, métadonnées) : des documents retrouvés et analysés plus facilement
- ⊕ Des millions de textes et de métadonnées téléchargeables facilement
- ⊕ Des connexions vers des outils / plateformes du monde académique

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources
- ⊕ Des données interopérables, des formats homogénéisés et des données corrigées (moins de prétraitements)
- ⊕ Des données enrichies (réocérisation, structuration de texte, métadonnées) : des documents retrouvés et analysés plus facilement
- ⊕ Des millions de textes et de métadonnées téléchargeables facilement
- ⊕ Des connexions vers des outils / plateformes du monde académique
- ⊕ Un cadre juridique sécurisé par une licence appropriée et déjà négociée

Présentation d'Istex

ISTEX DANS LE PAYSAGE DES RÉSERVOIRS | BILAN

Istex : le TDM en toute indépendance

- ⊕ Des données vérifiées accessibles un seul lieu pour de nombreuses sources
- ⊕ Des données interopérables, des formats homogénéisés et des données corrigées (moins de prétraitements)
- ⊕ Des données enrichies (réocérisation, structuration de texte, métadonnées) : des documents retrouvés et analysés plus facilement
- ⊕ Des millions de textes et de métadonnées téléchargeables facilement
- ⊕ Des connexions vers des outils / plateformes du monde académique
- ⊕ Un cadre juridique sécurisé par une licence appropriée et déjà négociée
- ⊕ Un accompagnement personnalisé avec **une équipe incroyable !**

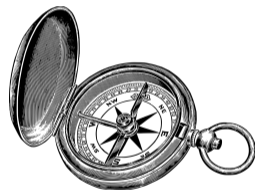


2. Constitution d'un corpus spécialisé

Constitution d'un corpus spécialisé

CAS D'USAGE

“Je suis doctorante en sociolinguistique, je débute une thèse sur les discours scientifiques à propos des réfugié-es dans les SHS au XXIe siècle. Je souhaite construire un corpus, distinguer la part de documents traitant des réfugié-es climatiques et des réfugié-es politiques et enfin identifier leur origine géographique.”



Constitution d'un corpus spécialisé

MÉTHODOLOGIE

1. Construire un corpus (cf. de Salabert et Barreaux, 2020)

- Vérifier la pertinence scientifique du corpus

Constitution d'un corpus spécialisé

MÉTHODOLOGIE

1. Construire un corpus (cf. de Salabert et Barreaux, 2020)

- Vérifier la pertinence scientifique du corpus
- Affiner ou enrichir le corpus au moyen d'outils propres à Istex en vue d'une exploitation TDM

Constitution d'un corpus spécialisé

MÉTHODOLOGIE

1. Construire un corpus (cf. de Salabert et Barreaux, 2020)

- Vérifier la pertinence scientifique du corpus
- Affiner ou enrichir le corpus au moyen d'outils propres à Istex en vue d'une exploitation TDM

2. Analyser les documents

- Employer des web services dans Lodex (extraction de thématiques, cartographie des pays cités, etc.)

Constitution d'un corpus spécialisé

MÉTHODOLOGIE

1. Construire un corpus (cf. de Salabert et Barreaux, 2020)

- Vérifier la pertinence scientifique du corpus
- Affiner ou enrichir le corpus au moyen d'outils propres à Istex en vue d'une exploitation TDM

2. Analyser les documents

- Employer des web services dans Lodex (extraction de thématiques, cartographie des pays cités, etc.)
- Catégoriser les documents : générer des clusters (GarganText, CorText, etc.)

Constitution d'un corpus spécialisé

MÉTHODOLOGIE

1. Construire un corpus (cf. de Salabert et Barreaux, 2020)

- Vérifier la pertinence scientifique du corpus
- Affiner ou enrichir le corpus au moyen d'outils propres à Istex en vue d'une exploitation TDM

2. Analyser les documents

- Employer des web services dans Lodex (extraction de thématiques, cartographie des pays cités, etc.)
- Catégoriser les documents : générer des clusters (GarganText, CorText, etc.)
- Faire des analyses lexicales, de la lexicométrie, de la statistique textuelle (Nooj, TXM, etc.)

Constitution d'un corpus spécialisé

MÉTHODOLOGIE

1. Construire un corpus (cf. de Salabert et Barreaux, 2020)

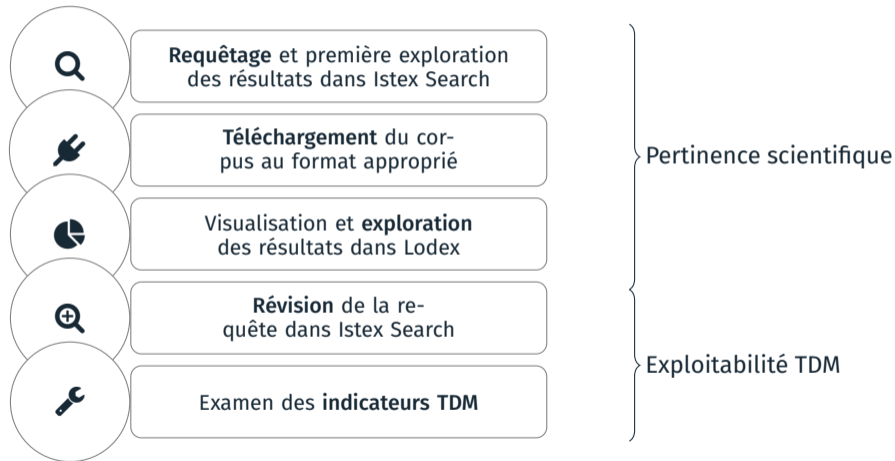
- Vérifier la pertinence scientifique du corpus
- Affiner ou enrichir le corpus au moyen d'outils propres à Istex en vue d'une exploitation TDM

2. Analyser les documents

- Employer des web services dans Lodex (extraction de thématiques, cartographie des pays cités, etc.)
- Catégoriser les documents : générer des clusters (GarganText, CorText, etc.)
- Faire des analyses lexicales, de la lexicométrie, de la statistique textuelle (Nooj, TXM, etc.)
- Utiliser et / ou adapter des outils d'analyse de sentiments (TextBlob, Pattern, etc.)

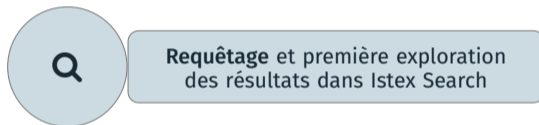
Constitution d'un corpus spécialisé

MÉTHODOLOGIE | 2 PHASES & 2 OUTILS



Constitution d'un corpus spécialisé

REQUÊTAGE



Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage

Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage
- **Consulter** un document dans une notice détaillée

Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage
- **Consulter** un document dans une notice détaillée
- **Naviguer** dans les résultats avec différents modes de tri et différents filtres

Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage
- **Consulter** un document dans une notice détaillée
- **Naviguer** dans les résultats avec différents modes de tri et différents filtres
- **Vérifier** la compatibilité TDM grâce à des indicateurs techniques

Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage
- **Consulter** un document dans une notice détaillée
- **Naviguer** dans les résultats avec différents modes de tri et différents filtres
- **Vérifier** la compatibilité TDM grâce à des indicateurs techniques
- **Sélectionner / exclure** des documents pour éliminer simplement le bruit

Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage
- **Consulter** un document dans une notice détaillée
- **Naviguer** dans les résultats avec différents modes de tri et différents filtres
- **Vérifier** la compatibilité TDM grâce à des indicateurs techniques
- **Sélectionner / exclure** des documents pour éliminer simplement le bruit
- **Télécharger** massivement des documents au format approprié

Constitution d'un corpus spécialisé

REQUÊTAGE | PRÉSENTATION DE SEARCH

Un seul outil pour

- **Interroger** l'API grâce à trois modes de requêtage
- **Consulter** un document dans une notice détaillée
- **Naviguer** dans les résultats avec différents modes de tri et différents filtres
- **Vérifier** la compatibilité TDM grâce à des indicateurs techniques
- **Sélectionner / exclure** des documents pour éliminer simplement le bruit
- **Télécharger** massivement des documents au format approprié

<https://search.istex.fr>

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 1

```
refugee OR réfugié OR asylum seeker OR  
demandeur d'asile OR migrant OR immigré  
OR immigrant
```

Description Liste de mots-clés anglais et français, coordonnés par l'opérateur booléen OR (= *ou inclusif*), insensibilité à la casse mais sensibilité aux diacritiques (chat = CHAT ; éclair ≠ eclair)

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 1

```
refugee OR réfugié OR asylum seeker OR  
demandeur d'asile OR migrant OR immigré  
OR immigrant
```

Description Liste de mots-clés anglais et français, coordonnés par l'opérateur booléen OR (= *ou inclusif*), insensibilité à la casse mais sensibilité aux diacritiques (chat = CHAT ; éclair ≠ eclair)

Résultats 412 299 documents

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 1

```
refugee OR réfugié OR asylum seeker OR  
demandeur d'asile OR migrant OR immigré  
OR immigrant
```

Description Liste de mots-clés anglais et français, coordonnés par l'opérateur booléen OR (= *ou inclusif*), insensibilité à la casse mais sensibilité aux diacritiques (chat = CHAT ; éclair ≠ eclair)

Résultats 412 299 documents

Problèmes En syntaxe Lucene le blanc équivaut à un OR, les expressions polylexicales doivent être entourées de guillemets ("demandeur d'asile")

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 2

```
refugee réfugié "asylum seeker" "demandeur  
d'asile" migrant immigré immigrant
```

Description Liste de mots-clés anglais et français, coordonnés par l'opérateur booléen OR, insensibilité à la casse mais sensibilité aux diacritiques

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 2

```
refugee réfugié "asylum seeker" "demandeur  
d'asile" migrant immigré immigrant
```

Description Liste de mots-clés anglais et français, coordonnés par l'opérateur booléen OR, insensibilité à la casse mais sensibilité aux diacritiques

Résultats 282 504 documents

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 2

```
refugee réfugié "asylum seeker" "demandeur  
d'asile" migrant immigré immigrant
```

Description Liste de mots-clés anglais et français, coordonnés par l'opérateur booléen OR, insensibilité à la casse mais sensibilité aux diacritiques

Résultats 282 504 documents

Problèmes Il y a du silence car les tokens ne sont pas lemmatisés : e.g. migrante, immigrées

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 3

```
refugee refugees réfugié réfugiée réfugiés  
réfugiées "asylum seeker" "asylum seekers"  
"demandeur d'asile" "demandeurs d'asile"  
"demandeuse d'asile" "demandeuses d'asile"  
migrant migrants migrante migrantes immigré  
immigrés immigrée immigrées immigrant  
immigrante immigrants immigrantes
```

Description Recherche des variantes morphosyntaxiques (féminins, pluriels)

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 3

```
refugee refugees réfugié réfugiée réfugiés  
réfugiées "asylum seeker" "asylum seekers"  
"demandeur d'asile" "demandeurs d'asile"  
"demandeuse d'asile" "demandeuses d'asile"  
migrant migrants migrante migrantes immigré  
immigrés immigrée immigrées immigrant  
immigrante immigrants immigrantes
```

Description Recherche des variantes morphosyntaxiques (féminins, pluriels)

Résultats 473 910 documents

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 3

```
refugee refugees réfugié réfugiée réfugiés  
réfugiées "asylum seeker" "asylum seekers"  
"demandeur d'asile" "demandeurs d'asile"  
"demandeuse d'asile" "demandeuses d'asile"  
migrant migrants migrante migrantes immigré  
immigrés immigrée immigrées immigrant  
immigrante immigrants immigrantes
```

Description Recherche des variantes morphosyntaxiques (féminins, pluriels)

Résultats 473 910 documents

Problèmes Il y a du silence : e.g. possessif anglais *refugee's*, écriture inclusive *réfugié·es*, formes dérivées *refugeeism*

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 4

```
refugee* réfugié* "asylum seeker" "asylum  
seekers" "demandeur d'asile" "demandeurs  
d'asile" "demandeuse d'asile" "demandeuses  
d'asile" migra* immigré* immigra*
```

Description Recherche des formes morphologiquement apparentées grâce à la **troncature** symbolisée par *

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 4

```
refugee* réfugié* "asylum seeker" "asylum  
seekers" "demandeur d'asile" "demandeurs  
d'asile" "demandeuse d'asile" "demandeuses  
d'asile" migra* immigré* immigra*
```

Description Recherche des formes
morphologiquement apparentées
grâce à la **troncature** symbolisée
par *

Résultats 2 286 615 documents

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 4

```
refugee* réfugié* "asylum seeker" "asylum  
seekers" "demandeur d'asile" "demandeurs  
d'asile" "demandeuse d'asile" "demandeuses  
d'asile" migra* immigré* immigra*
```

Description Recherche des formes
morphologiquement apparentées
grâce à la **troncature** symbolisée
par *

Résultats 2 286 615 documents

Problèmes Il y a du bruit

Constitution d'un corpus spécialisé

REQUÊTAGE

Introduction

The Albian vegetation of western Europe has been studied in several works which described micro-, meso- and macroremains [1–7], and highlighting the prevalence of ferns and gymnosperms—mainly conifers (Araucariaceae-Podocarpaceae-Pinaceae-Cupressaceae-Cheirolepidiaceae)—, and the increasing representation of angiosperms. The distinctiveness of the western European flora has been recognised for a long time and several biogeographic provinces have been named on the basis of the nature of the fossil assemblages (i.e., palynological and megaremain records [8, 9]). The increased attention to the time-interval is justified by the rising of the dominance of the flowering plants in many ecosystems [10–13] including arid environments [14]. During the later part of the Early Cretaceous, SW Europe has been described as one of the probable migratory routes of angiosperms originating from a northern Gondwanan centre of radiation [15]. The successive northward migrations of angiosperms, well-characterised by both pollen and megaremain records, suggest broadly similar climate conditions in low to mid-latitude settings (i.e., low latitudinal climatic gradient) or selective migration of taxa able to tolerate a wide range of climatic conditions [16]. During most of the Albian, Iberia was part of a climatic zone referred to as the Northern Hot Arid (NHA) belt [17], where a desert system (erg) developed over an extensive area [18].

Constitution d'un corpus spécialisé

REQUÊTAGE

Introduction

The Albian vegetation of western Europe has been studied in several works which described micro-, meso- and macroremains [1–7], and highlighting the prevalence of ferns and gymnosperms—mainly conifers (Araucariaceae-Podocarpaceae-Pinaceae-Cupressaceae-Cheirolepidiaceae)—, and the increasing representation of angiosperms. The distinctiveness of the western European flora has been recognised for a long time and several biogeographic provinces have been named on the basis of the nature of the fossil assemblages (i.e., palynological and megaremain records [8, 9]). The increased attention to the time-interval is justified by the rising of the dominance of the flowering plants in many ecosystems [10–13] including arid environments [14]. During the later part of the Early Cretaceous, SW Europe has been described as one of the probable migratory routes of angiosperms originating from a northern Gondwanan centre of radiation [15]. The successive northward migrations of angiosperms, well-characterised by both pollen and megaremain records, suggest broadly similar climate conditions in low to mid-latitude settings (i.e., low latitudinal climatic gradient) or selective migration of taxa able to tolerate a wide range of climatic conditions [16]. During most of the Albian, Iberia was part of a climatic zone referred to as the Northern Hot Arid (NHA) belt [17], where a desert system (erg) developed over an extensive area [18].

Bull. Fr. Pêche Piscic. (1995) 337/338/339 : 47-61

— 47 —

L'HABITAT À UNE ÉCHELLE GLOBALE : IMPORTANCE RESPECTIVE DE L'IMMIGRATION ET DE L'EXTINCTION POUR LES PEUPELEMENTS DE POISSONS DES PETITS LACS.

W.M. TONN, R.E. VANDENBOS et C.A. PASZKOWSKI

Department of Biological Sciences, University of Alberta,
Edmonton, Alberta T6G 2E9, Canada.

RÉSUMÉ (traduit par les éditeurs)

Les deux processus qui déterminent la structure des communautés dans les habitats insulaires sont l'immigration d'espèces vers une communauté et l'extinction de celles qui étaient présentes auparavant. Bien que l'immigration et l'extinction se produisent généralement à des échelles supérieures à celles qu'il est possible d'étudier avec des

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 5

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

Description Exclusion des termes polysémiques, sélection des termes consacrés (droit international)

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 5

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

Description Exclusion des termes polysémiques, sélection des termes consacrés (droit international)

Résultats 140 062 documents

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 5

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

Description Exclusion des termes polysémiques, sélection des termes consacrés (droit international)

Résultats 140 062 documents

Problèmes Il y a du bruit (*où sont les termes détectés ?*)

Constitution d'un corpus spécialisé

REQUÊTAGE

11. Bankole A, Biddlecom A, Guiella G, Singh S, Zulu E. Sexual behavior, knowledge and information sources of very young adolescents in four sub-Saharan African countries. *Afr J Reprod Health* [Internet]. 2007; 11(3):28–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18458739%5Chttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2367131> PMID: 18458739
12. Atuyambe LM, Kibira SPS, Bukenya J, Muhumuza C, Apolot RR, Mulogo E. Understanding sexual and reproductive health needs of adolescents: evidence from a formative evaluation in Wakiso district, Uganda. *Reprod Health* [Internet]. 2015; 12:35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25896066%5Chttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4416389> <https://doi.org/10.1186/s12978-015-0026-7> PMID: 25896066
13. Seifu A, Fantahun M, Worku A. Reproductive health needs of out-of-school adolescents: A cross-sectional comparative study of rural and urban areas in northwest Ethiopia. *JHealth Dev* [Internet]. 2006; 20(1):10–7. Available from: <https://pdfs.semanticscholar.org/02f9/e464ddfc323954e7d1b7600f0bd3641bc862.pdf>
14. Ortiz-Echevarria L, Greeley M, Bawoke T, Zimmerman L, Robinson C, Schlecht J. Understanding the unique experiences, perspectives and sexual and reproductive health needs of very young adolescents: Somali refugees in Ethiopia. *Confl Health*. 2017; 11(Suppl 1).

Constitution d'un corpus spécialisé

REQUÊTAGE

11. Bankole A, Biddlecom A, Guiella G, Singh S, Zulu E. Sexual behavior, knowledge and information sources of very young adolescents in four sub-Saharan African countries. *Afr J Reprod Health* [Internet]. 2007; 11(3):28–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18458739> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC2367131> PMID: 18458739
12. Atuyambe LM, Kibira SPS, Bukenya J, Muhumuza C, Apolot RR, Mulogo E. Understanding sexual and reproductive health needs of adolescents: evidence from a formative evaluation in Wakiso district, Uganda. *Reprod Health* [Internet]. 2015; 12:35. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25896066> <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4416389> <https://doi.org/10.1186/s12978-015-0026-7> PMID: 25896066
13. Seifu A, Fantahun M, Worku A. Reproductive health needs of out-of-school adolescents: A cross-sectional comparative study of rural and urban areas in northwest Ethiopia. *JHealth Dev* [Internet]. 2006; 20(1):10–7. Available from: <https://pdfs.semanticscholar.org/02f9/e464ddfc323954e7d1b7600f0bd3641bc862.pdf>
14. Ortiz-Echevarria L, Greeley M, Bawoke T, Zimmerman L, Robinson C, Schlecht J. Understanding the unique experiences, perspectives and sexual and reproductive health needs of very young adolescents: Somali refugees in Ethiopia. *Confl Health*. 2017; 11(Suppl 1).

Douleurs musculo-squelettiques non systématisées diffuses et déficit sévère en vitamine D. Étude avant-après en médecine générale

u, Edwige Bodier, Gilbert Souweine, Alain Moreau, Corinne istian Dupraz

Université médicale Lyon Est, 8, avenue Rockefeller, 69008 Lyon, France

Correspondance :

Marie France Le Goaziou, Université médicale Lyon Est, 8, avenue Rockefeller, 69008 Lyon, France.

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 6

```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")))
```

Description Cibler les champs Istex les plus à même de renvoyer des documents pertinents : title, abstract, subject.value, keywords.teeft, exclure les documents qui contiennent uniquement les termes dans les refBibs.title grâce au parenthésage

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 6

```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile"))
```

Description Cibler les champs Istex les plus à même de renvoyer des documents pertinents : title, abstract, subject.value, keywords.teeft, exclure les documents qui contiennent uniquement les termes dans les refBibs.title grâce au **parenthésage**

Résultats 26 929 documents

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 6

```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile"))
```

Description Cibler les champs Istex les plus à même de renvoyer des documents pertinents : title, abstract, subject.value, keywords.teeft, exclure les documents qui contiennent uniquement les termes dans les refBibs.title grâce au **parenthésage**

Résultats 26 929 documents

Problèmes Il y a du bruit (*objectif scientifique*)

Constitution d'un corpus spécialisé

REQUÊTAGE

CATÉGORIE WOS (189) ▾

CATÉGORIE SCIENCE-METRIX (178) ▾

CATÉGORIE SCOPUS (309) ▲

Rechercher 🔍

<input type="checkbox"/> 1 - Social Sciences	16 486
<input type="checkbox"/> 2 - Social Sciences	14 157
<input type="checkbox"/> 1 - Health Sciences	6 355
<input type="checkbox"/> 2 - Arts and Humanities	6 007
<input type="checkbox"/> 2 - Medicine	5 969

APPLIQUER ✕

CATÉGORIE INIST (72) ▾

Constitution d'un corpus spécialisé

REQUÊTAGE

CATÉGORIE WOS (189) ▼

CATÉGORIE SCIENCE-METRIX (178) ▼

CATÉGORIE SCOPUS (309) ▲

Rechercher

▲▼

<input type="checkbox"/> 1 - Social Sciences	16 486
<input type="checkbox"/> 2 - Social Sciences	14 157
<input type="checkbox"/> 1 - Health Sciences	6 355
<input type="checkbox"/> 2 - Arts and Humanities	6 007
<input type="checkbox"/> 2 - Medicine	5 969

CATÉGORIE INIST (72) ▼

DATE DE PUBLICATION ▲

Année (1595 à 2023)

à

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 7

```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")) AND categories.scopus:
("1 - Social Sciences") AND publicationDate:
[2000 TO *]
```

Description Sélectionner les catégories scientifiques grâce à l'opérateur booléen AND, cibler les documents du XXIe siècle grâce à l'intervalle [2000 TO *]

Constitution d'un corpus spécialisé

REQUÊTAGE

Requête 7

```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")) AND categories.scopus:
("1 - Social Sciences") AND publicationDate:
[2000 TO *]
```

Description Sélectionner les catégories scientifiques grâce à l'opérateur booléen AND, cibler les documents du XXIe siècle grâce à l'intervalle [2000 TO *]

Résultats 7 020 documents

Constitution d'un corpus spécialisé

REQUÊTE

Requête 8

```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")) AND categories.scopus:
("1 - Social Sciences") AND publicationDate:
[2000 TO *] AND abstract:*
```

Description Cibler les documents avec un résumé grâce à `abstract:*`

Constitution d'un corpus spécialisé

REQUÊTE

Requête 8

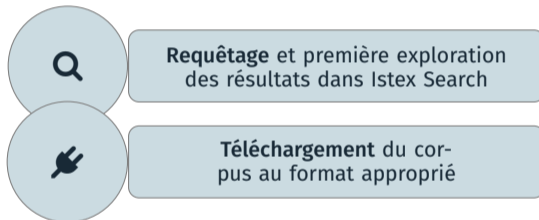
```
(title:(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") abstract:(refugee*
réfugié* "asylum seeker" "asylum seekers"
"demandeur d'asile" "demandeurs d'asile"
"demandeuse d'asile" "demandeuses d'asile")
subject.value:(refugee* réfugié* "asylum
seeker" "asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")(keywords.teeft:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile") NOT refBibs.title:
(refugee* réfugié* "asylum seeker"
"asylum seekers" "demandeur d'asile"
"demandeurs d'asile" "demandeuse d'asile"
"demandeuses d'asile")) AND categories.scopus:
("1 - Social Sciences") AND publicationDate:
[2000 TO *] AND abstract:*
```

Description Cibler les documents avec un résumé grâce à `abstract:*`

Résultats 4 866 documents

Constitution d'un corpus spécialisé

TÉLÉCHARGEMENT



Constitution d'un corpus spécialisé

TÉLÉCHARGEMENT

Configurez votre téléchargement

USAGE PERSONNALISÉ **LODEX** CORTEXT GARGANTEXT

<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"/> Annexes	<input type="checkbox"/> refBibs
<input type="checkbox"/> ZIP	<input type="checkbox"/> Couvertures	<input type="checkbox"/> teeft
<input type="checkbox"/> TIFF		<input type="checkbox"/> unitex

Trier par : **date de publication** ▾ ↓

Télécharger / 4 866

Format de l'archive : **ZIP** ▾ Compression : **moyenne** ▾

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* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") abstract:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile"...
```

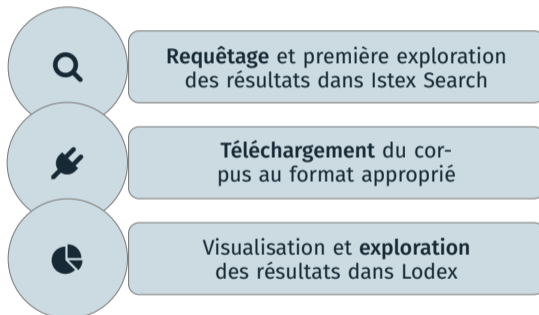
Requête brute complète

```
https://ap.Istex.fr/document?q=(title:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile"...
```

Modale de téléchargement des données dans Istex Search

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



« *Linked Open Data EXperiment* »

Application web **open-source** dédiée à la **visualisation** de données structurées

« *Linked Open Data EXperiment* »

Application web **open-source** dédiée à la **visualisation** de données structurées

<https://www.lodex.fr>

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web À partir de différents formats (CSV, JSON, TXT, ...)

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web

À partir de différents formats (CSV, JSON, TXT, ...)



Explorer ses données

À l'aide de différents graphiques, filtres et au travers de données complémentaires

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web

À partir de différents formats (CSV, JSON, TXT, ...)



Explorer ses données

À l'aide de différents graphiques, filtres et au travers de données complémentaires



Enrichir ses données

Grâce à un catalogue de [web services](#)

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web À partir de différents formats (CSV, JSON, TXT, ...)



Explorer ses données À l'aide de différents graphiques, filtres et au travers de données complémentaires



Enrichir ses données Grâce à un catalogue de [web services](#)



Aligner ses données Avec des données similaires ou connexes

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web À partir de différents formats (CSV, JSON, TXT, ...)



Explorer ses données À l'aide de différents graphiques, filtres et au travers de données complémentaires



Enrichir ses données Grâce à un catalogue de [web services](#)



Aligner ses données Avec des données similaires ou connexes



Exporter ses données En format classique ou du web sémantique

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web À partir de différents formats (CSV, JSON, TXT, ...)



Explorer ses données À l'aide de différents graphiques, filtres et au travers de données complémentaires



Enrichir ses données Grâce à un catalogue de [web services](#)



Aligner ses données Avec des données similaires ou connexes



Exporter ses données En format classique ou du web sémantique



Attribuer des identifiants ARK, URI

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Transformer ses données en site web À partir de différents formats (CSV, JSON, TXT, ...)



Explorer ses données À l'aide de différents graphiques, filtres et au travers de données complémentaires



Enrichir ses données Grâce à un catalogue de [web services](#)



Aligner ses données Avec des données similaires ou connexes



Exporter ses données En format classique ou du web sémantique

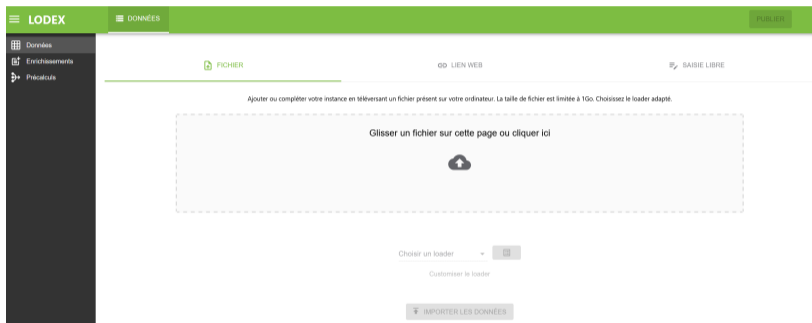


Attribuer des identifiants ARK, URI

Exemple

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



Interface d'import des données dans Lodex

Constitution d'un corpus spécialisé

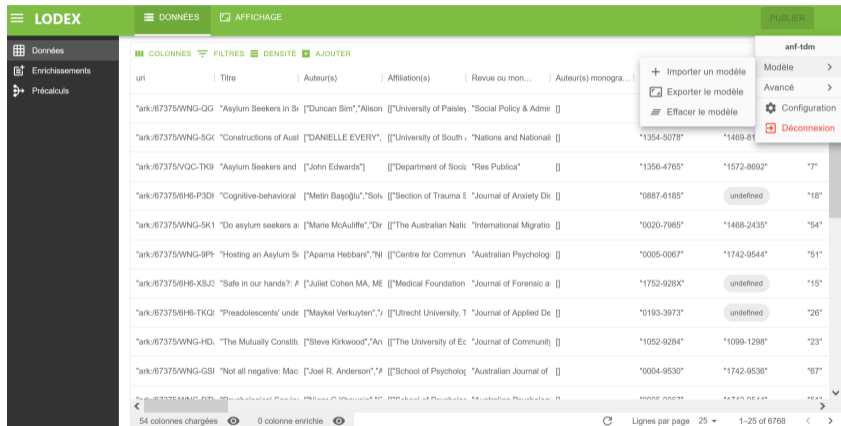
EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX

uri	Titre	Auteur(s)	Affiliation(s)	Revue ou mon...	Auteur(s) monogra...	ISSN	e-ISSN	Volume
"ark:/67375/WNG-QG	"Asylum Seekers in S	["Duncan Sim", "Alison	["University of Paisley,	"Social Policy & Admir		"0144-5596"	"1467-9515"	"41"
"ark:/67375/WNG-5G	"Constructions of Aust	["DANIELLE EVERY",	["University of South,	"Nations and Nationali		"1354-5078"	"1469-8129"	"14"
"ark:/67375/VQC-TK9	"Asylum Seekers and	["John Edwards"]	["Department of Soci	"Res Publica"		"1356-4765"	"1572-8692"	"7"
"ark:/67375/6H6-P3D	"Cognitive-behavioral	["Metin Başoğlu", "Solv	["Section of Trauma &	"Journal of Anxiety Dis		"0887-6185"	undefined	"18"
"ark:/67375/WNG-5K1	"Do asylum seekers a	["Marie McAuliffe", "Dir	["The Australian Natic	"International Migratio		"0020-7985"	"1468-2435"	"54"
"ark:/67375/WNG-9P	"Hosting an Asylum S	["Aparna Hebbani", "Ni	["Centre for Commun	"Australian Psychologi		"0005-0067"	"1742-9544"	"51"
"ark:/67375/6H6-XSJ	"Safe in our hands?: A	["Juliet Cohen MA, ME	["Medical Foundation	"Journal of Forensic a		"1752-928X"	undefined	"15"
"ark:/67375/6H6-TKQ	"Preadolescents' unde	["Maykel Verkuyten", "v	["Utrecht University, T	"Journal of Applied De		"0193-3973"	undefined	"26"
"ark:/67375/WNG-HD	"The Mutually Constit.	["Steve Kirkwood", "An	["The University of Ec	"Journal of Community		"1052-9284"	"1099-1298"	"23"
"ark:/67375/WNG-GSI	"Not all negative: Mac	["Joel R. Anderson", "A	["School of Psycholog	"Australian Journal of		"0004-9530"	"1742-9536"	"67"

Transformation des données par le loader

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | PRÉSENTATION DE LODEX



The screenshot displays the LODEX application interface. At the top, there is a green navigation bar with the LODEX logo and tabs for 'DONNÉES' and 'AFFICHAGE'. A 'PUBLIER' button is visible on the right. On the left, a dark sidebar contains navigation options: 'Données', 'Enrichissements', and 'Précalculs'. The main area shows a table with columns for 'uri', 'Titre', 'Auteur(s)', 'Affiliation(s)', 'Revue ou mon...', and 'Auteur(s) monogra...'. A context menu is open over the table, offering options: 'Importer un modèle', 'Exporter le modèle', and 'Effacer le modèle'. To the right of the table, there is a 'publié' column with a dropdown menu showing 'anf-tdm', 'Modèle', 'Avancé', 'Configuration', and 'Déconnexion'. At the bottom of the table, it indicates '54 colonnes chargées' and '0 colonne enrichie'. The footer shows 'Lignes par page 25' and '1-25 of 6768'.

uri	Titre	Auteur(s)	Affiliation(s)	Revue ou mon...	Auteur(s) monogra...	publié
*ark:/67375/WNG-QG	"Asylum Seekers in St	["Duncan Sim", "Alison	["University of Paisley	"Social Policy & Admir		
*ark:/67375/WNG-5GK	"Constructions of Aust	["DANIELLE EVERY",	["University of South	"Nations and Nationali		*1354-5078* *1469-81
*ark:/67375/VQC-TK9	"Asylum Seekers and	["John Edwards"]	["Department of Soci	"Res Publica"		*1356-4785* *1572-8692* *7*
*ark:/67375/6H6-P3DI	"Cognitive-behavioral	["Melin Başoğlu", "Solv	["Section of Trauma &	"Journal of Anxiety Dis		*0887-6185* undefined *18*
*ark:/67375/WNG-5K1	"Do asylum seekers ai	["Marie McAuliffe", "Dir	["The Australian Natic	"International Migratio		*0020-7985* *1468-2435* *54*
*ark:/67375/WNG-9Ph	"Hosting an Asylum Si	["Aparna Hebbani", "NI	["Centre for Commun	"Australian Psychologi		*0005-0067* *1742-9544* *51*
*ark:/67375/6H6-XSJC	"Safe in our hands?: A	["Juliet Cohen MA, ME	["Medical Foundation	"Journal of Forensic a		*1752-928X* undefined *15*
*ark:/67375/6H6-TKQI	"Preadolescents' unde	["Maykel Verkuuten", "y	["Utrecht University, T	"Journal of Applied De		*0193-3973* undefined *26*
*ark:/67375/WNG-HD.	"The Mutually Consti.	["Steve Kirkwood", "An	["The University of Ec	"Journal of Communit		*1052-9284* *1099-1298* *23*
*ark:/67375/WNG-GSI	"Not all negative: Mac	["Joel R. Anderson", "A	["School of Psycholo	"Australian Journal of		*0004-9530* *1742-9536* *67*

Import du modèle

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION | SITE DU CORPUS

ISTEX Services

Les technologies et les outils ISTEX pour les projets de recherche

Réfugiés et demandeurs d'asile



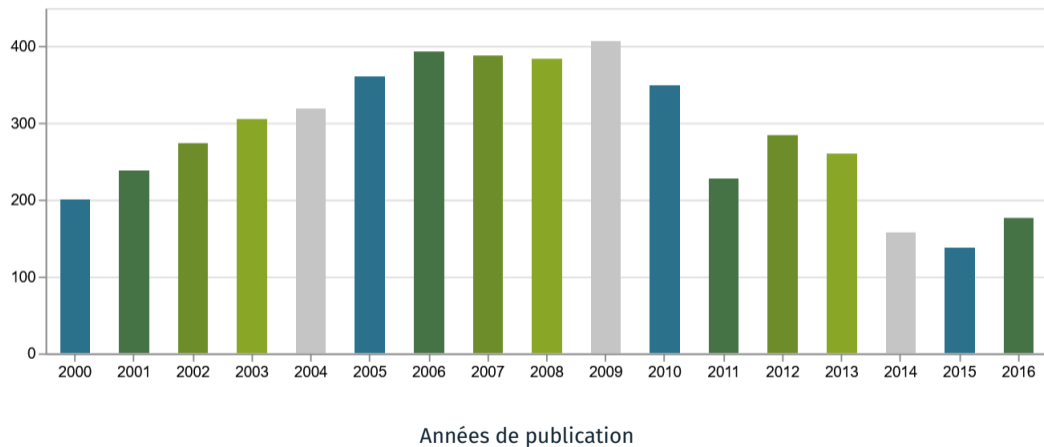
Image par [Ciker-Free-Vector-Images](#) de Pixabay.

Le corpus *Réfugiés et demandeurs d'asile* contient des publications [Istex](#) relatives à **l'accueil des réfugiés et demandeurs d'asile** à travers le monde, leur situation sanitaire et sociale et leur qualité de vie, ainsi que l'attitude politique et le positionnement juridique de différents pays sur cette question. Son but est de permettre l'analyse des discours à propos des réfugiés dans les SHS, de distinguer la part de documents traitant des réfugiés climatiques et des réfugiés politiques et d'identifier leur origine géographique.

Site web du corpus créé avec Lodex

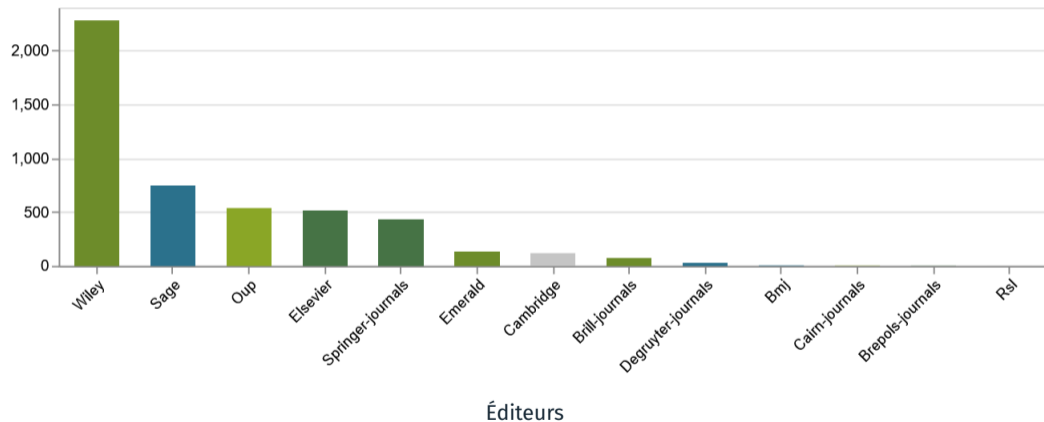
Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



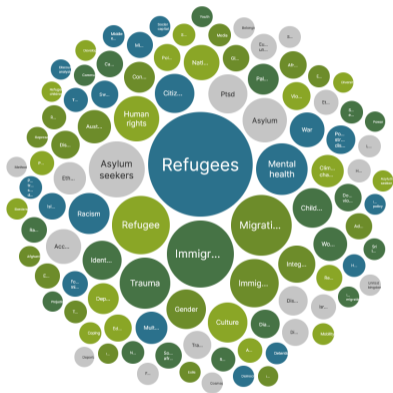
Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



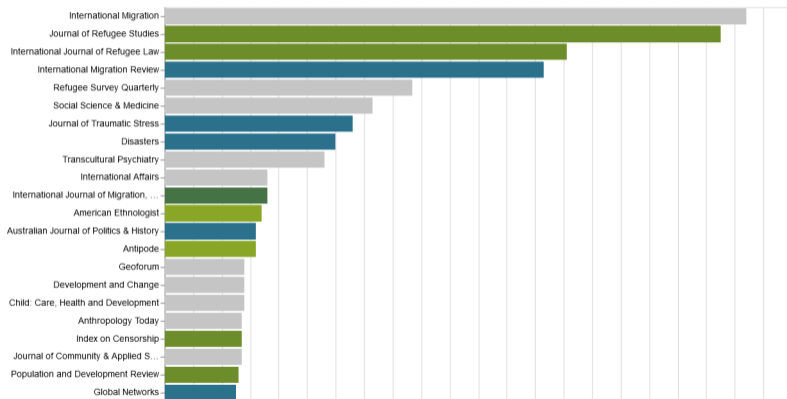
Mots-clés d'auteurs



Mots-clés Teft

Constitution d'un corpus spécialisé

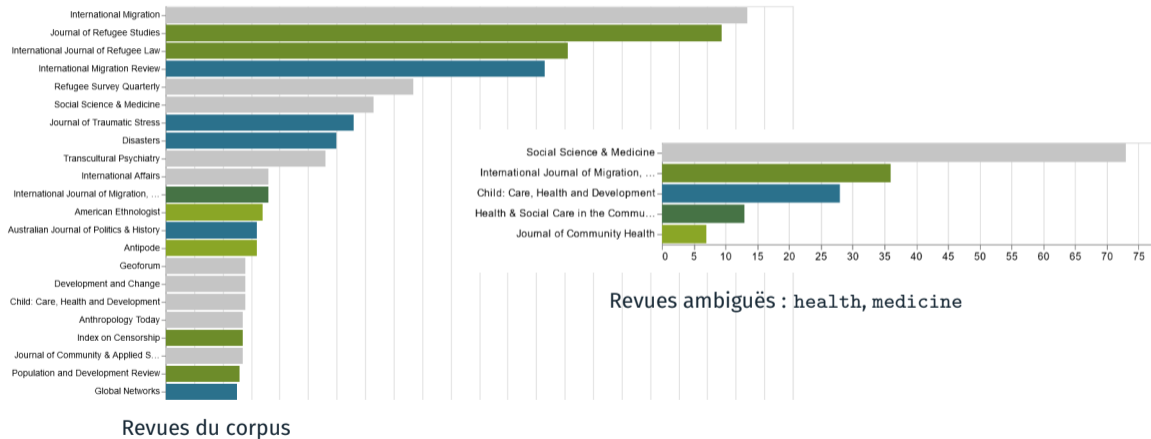
EXPLORATION ET VISUALISATION



Reuves du corpus

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION

J Community Health (2009) 34:173–180
DOI 10.1007/s10900-008-9141-5

ORIGINAL PAPER

Hepatitis B Testing and Vaccination Among Vietnamese- and Cambodian-Americans

Scott P. Grytdal · Youlian Liao · Roxana Chen · Cheza C. Garvin ·
Dorcas Grigg-Saito · Marjorie Kagawa-Singer · Sidney Liang ·
Stephen J. McPhee · Tung T. Nguyen · Jacqueline H. Tran · Kathleen M. Gallagher

Published online: 21 February 2009
© Centers for Disease Control and Prevention 2009

Abstract We determined hepatitis B virus (HBV) testing and vaccination levels and factors associated with testing and vaccination among Vietnamese- and Cambodian-Americans. We also examined factors associated with healthcare professional (HCP)-patient discussions about HBV. We analyzed 2006 Racial and Ethnic Approaches to Community Health (REACH) 2010 Risk Factor Survey data from four US communities. We used logistic regression to identify variables associated with HBV vaccination, testing, and HCP-patient discussions about HBV. Of the 2,049 Vietnamese- and Cambodian-American respondents, 60% reported being tested for HBV, 35% reported being vaccinated against hepatitis B, and 36% indicated that they had discussed HBV with a HCP. Cambodian-Americans were less likely than Vietnamese-Americans to have been tested for HBV, while respondents with at least a high school diploma were more likely to have been tested for HBV. Respondents born in the US, younger individuals, and

respondents with at least some college education were more likely to have been vaccinated against hepatitis B. HBV testing and vaccination remain suboptimal among members of these populations. Culturally sensitive efforts that target Vietnamese- and Cambodian-Americans for HBV testing and vaccination are needed to identify chronic carriers of HBV, prevent new infections, and provide appropriate medical management. HCPs that serve these populations should be encouraged to discuss HBV with their patients.

Keywords Hepatitis B · Screening · Vaccination · Refugee/immigrant health

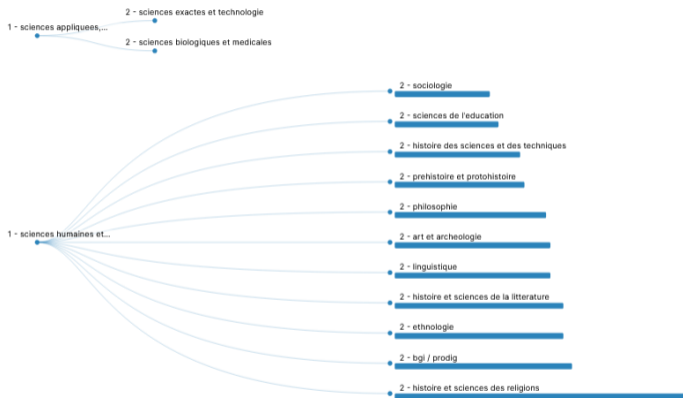
Introduction

Globally, chronic hepatitis B virus (HBV) infection is an important public health concern. Persons with chronic HBV infection are at increased risk of developing cirrhosis,

Exemple d'article provoquant du bruit issu de la revue Journal of Community Health

Constitution d'un corpus spécialisé

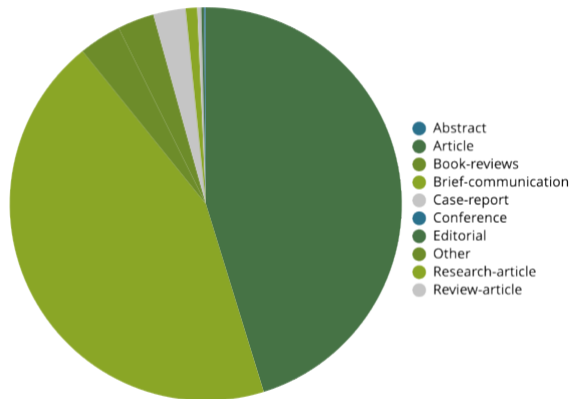
EXPLORATION ET VISUALISATION



Catégories Inist correspondantes : 1 - sciences appliquées, technologies et médecine

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



Types de documents

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION

Language learning

by the administrative people involved. This paper presents a case study of the implementation of an EFL innovation in the form of CALL (Computer Assisted Language Learning) in a Turkish State University's EFL Preparatory School. It is argued that teachers should be engaged participants in the change process and that this increases the chance of successful implementation of the innovation concerned.

<http://www.elj.oxfordjournals.org>

07-197 WARD, MONICA (Dublin City U, Ireland),
Using software design methods in CALL.

Computer Assisted Language Learning
(Routledge/Taylor & Francis) 19.2-3 (2006),
129-147.
doi:10.1080/09588220600821487

The phrase 'software design' is not one that arouses the interest of many CALL practitioners, particularly those from a humanities background. However, software design essentials are simply logical ways of going about designing a system. The fundamentals include modularity, anticipation of change, generality and an incremental approach. While CALL researchers and developers might consider software design as something that is required for large-scale projects, it is relevant and useful for all types of projects. Colpaert's CALL design model incorporates software design principles and has a particular focus on CALL. This article reviews software design principles and their relevance in CALL. It focuses on how Colpaert's model can be applied in a real-world situation. One key feature is the importance of involving users (and especially teachers in the CALL context) in the design process. Although this can be challenging for the designer and the design-novice teacher, it is an essential component in successful CALL projects.

differences. *ELT Journal* (Oxford University Press) 60.4 (2006), 365-373.
doi:10.1093/elt/ccl027

One of the most controversial issues in FL teaching is the age at which language learning should start. Nowadays it is recognized that in second language contexts maturational constraints make an early start advisable, but there is still disagreement regarding the problem of when to start or the best way to learn in foreign contexts. The aim of this paper is threefold: to establish if there is a critical or sensitive period for FL learners; to determine the particular linguistic and cognitive aspects affected by this period; and to make a pedagogical proposal to overcome the age-related problem using an extract taken from the film *Shrek*. This proposal comprises two lesson plans using the same film extract, one for children and the other one for older students. These plans are then compared in terms of the different cognitive, linguistic, and metalinguistic processes involved in learning and teaching for each age range.

<http://www.elj.oxfordjournals.org>

07-199 AMMAR, AHLEM (U de Montréal, Canada; ahlem.ammam@umontreal.ca) & **NINA SPADA, One size fits all? Recasts, prompts, and L2 learning.**
Studies in Second Language Acquisition
(Cambridge University Press) 28.4 (2006), 543-574.
doi:10.1017/S0272263106060268

This quasi-experimental study investigated the potential benefits of two corrective feedback techniques (recasts and prompts) for learners of different proficiency levels. Sixty-four students in three intact grade 6 intensive English as a second language classes in the Montreal area were assigned to the two experimental conditions— one received corrective feedback in the form of recasts

Bruit associé au type abstract (15 documents)

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION

Some men in some books is scarce, unfortunately, by many books on one subject—namely, women and gender are scarcely mentioned in either book, aside from a few passing remarks by Tibi noting that fundamentalists often address only men in their nouns and pronouns, effectively eliminating women from consideration. Although Hunter's book is mainly about foreign relations, it includes discussions of broader questions about Islamic theory and practice, and might also note that problems regarding foreign relations are influenced, at least on the U.S. side, by the treatment of women in some Islamic countries. Women, who are all too often omitted from discussion about political questions, are absolutely central to Islamist ideology and practice. The Islamist vision of Islam includes a return to universal veiling, however changed the endorsed costume may be, and to various degrees of sex segregation. In addition, Islamist pressure or rule in various Muslim countries has resulted in reverses in legal gains for women and greater tolerance for various forms of violence against, and mistreatment of, women. Women and gender are core-central in Islamist ideologies and practices, and to ignore this or play it down, as too many scholars do, leaves out some of the deepest roots of the fundamentalist appeal and also of opposition to Islamism. While some women have been able, in Iran and elsewhere, to organize effective "Islamic feminist" attacks on male supremacist ideas and practices, they have nowhere been able to establish, in either ideas or practices, anything like the equality in treatment and opportunity for women that many, including both Tibi and Hunter, think is needed.¹

With these caveats this reviewer welcomes two very good books to the growing, but not always equally enlightening, literature on Islamism or Islamic Fundamentalism, as you prefer.

NOTE

¹Points of entry into such discussions include several articles in the Winter 1999 special issue of the *Journal of Women's History, Women and Twentieth Century Religious Politics*, ed. Nikki R. Keddie and Jasmin Rostan-Kolaji. Debating Western scholarship on fundamentalism are Henry Munson, Jr., "Indelible Tolerance: Western Academia and Islamic Fundamentalism," and Beth Baron, "Tolerable hardness? Silence on Attacks on Women by Fundamentalists," *Contexts* 5, no. 3 (Spring 1996): 99–117, 119–26. These criticized in these articles declined to respond, but there was a response in the same issue by Mark Jeurgensmeijer, "Fundaphobia—The Irrational Fear of Fundamentalism," 27–132, and Munson, "Response to Jeurgensmeijer and Baron," 133–34.

ANDREW JARICHOW, *The Saudi File: People, Power, Politics* (New York: St. Martin's Press, 1998), Pp. 340. Price not available.

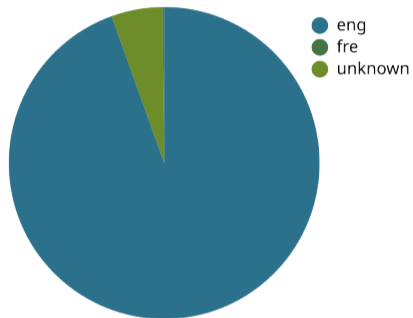
REVIEWED BY GWYNETH OGBURN, Department of Political Science, University of Arkansas, Fayetteville

The Saudi File is a handy volume, with limitations, for those who write on Saudi Arabia: it is a compilation of various documents, statements, press clips, and reports that broadly concern politics in the country. The chapter titles reflect important subjects, such as the Basic Law, government and *shura* council, political reform, an Islamic state, religious police, human rights, the press and freedom of speech, refugees, gender, labor, terror, security, international relations, dissidents abroad, and the Gulf War. The relevance of these issues and institutions makes the volume a convenient tool for research.

Bruit associé au type book-reviews (147 documents)

Constitution d'un corpus spécialisé

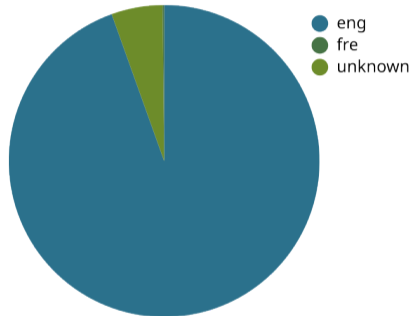
EXPLORATION ET VISUALISATION



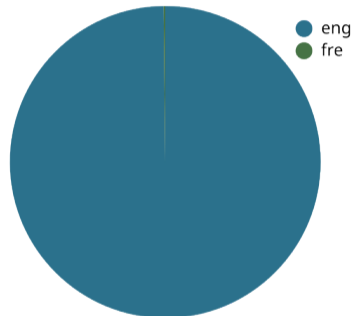
Langues : français (8), inconnue (259), anglais (4 599)

Constitution d'un corpus spécialisé

EXPLORATION ET VISUALISATION



Langues : français (8), inconnue (259), anglais (4 599)



Langues issues du WS

Constitution d'un corpus spécialisé

RÉVISION DE LA REQUÊTE



Constitution d'un corpus spécialisé

RÉVISION DE LA REQUÊTE | BILAN

Requête révisée

```
NOT categories.inist:"1 - sciences appliquees,  
technologies et medecines"
```

- **Domaines scientifiques** - Exclure les articles ambigus grâce aux catégories Inist

Constitution d'un corpus spécialisé

RÉVISION DE LA REQUÊTE | BILAN

Requête révisée

```
NOT categories.inist:"1 - sciences appliquees,  
technologies et medecines"
```

```
NOT genre:("book-reviews" "abstract")
```

- **Domaines scientifiques** - Exclure les articles ambigus grâce aux catégories Inist
- **Types de document** - Exclure les documents qui ne sont pas des articles et provoqueraient du bruit par la suite

Constitution d'un corpus spécialisé

RÉVISION DE LA REQUÊTE | BILAN

Requête révisée

```
NOT categories.inist:"1 - sciences appliquees,  
technologies et medecines"
```

```
NOT genre:("book-reviews" "abstract")
```

```
AND language:("eng" "fre")
```

- **Domaines scientifiques** - Exclure les articles ambigus grâce aux catégories Inist
- **Types de document** - Exclure les documents qui ne sont pas des articles et provoqueraient du bruit par la suite
- **Langues** - Le web service a montré que les documents associés aux langues inconnues étaient en anglais

Constitution d'un corpus spécialisé

RÉVISION DE LA REQUÊTE | BILAN

Requête révisée (-291 documents)

Résultats de votre requête

4 595 documents trouvés



RECHERCHER

```
((title:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile")abstract:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") subject.value:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile"))(keywords.teeft:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile") NOT refBibs.title:(refugee* réfugié* "asylum seeker" "asylum seekers" "demandeur d'asile" "demandeurs d'asile" "demandeuse d'asile" "demandeuses d'asile")))) AND categories.scopus:("1 - Social Sciences") AND publicationDate:[2000 TO *] AND abstract:* ) NOT categories.inist:"1 - sciences appliquees, technologies et medecines" NOT genre:( "book-reviews" "abstract")
```

Nouvelle requête

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un résumé

Constat 100% des documents ont un résumé

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un résumé

Constat 100% des documents ont un résumé

Outils GarganText : plateforme collaborative d'exploration graphique de corpus

Lodex : utilisation des résumés pour les web services

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un résumé

Constat 100% des documents ont un résumé

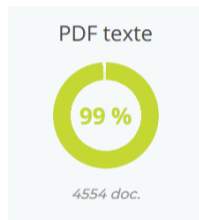
Outils GarganText : plateforme collaborative d'exploration graphique de corpus

Lodex : utilisation des résumés pour les web services

Requête AND abstract:*

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un PDF texte

Constat 99% des documents ont un PDF texte (e.g. pdf image)

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



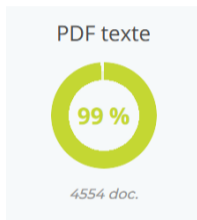
Pourcentage de documents possédant un PDF texte

Constat 99% des documents ont un PDF texte (e.g. pdf image)

Outils Transformer le PDF en texte avant l'intégration dans l'outil souhaité (*Tesseract*)

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un PDF texte

Constat 99% des documents ont un PDF texte (e.g. pdf image)

Outils Transformer le PDF en texte avant l'intégration dans l'outil souhaité (*Tesseract*)

Requête `qualityIndicators.pdfText:true`

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un texte nettoyé

Constat 75% des documents possèdent un **texte nettoyé**

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un texte nettoyé

Constat 75% des documents possèdent un **texte nettoyé**

Outils CorText : outil d'analyse et d'exploitation de corpus textuels

TXM : outil d'analyse textométrique de corpus de textes (Heiden *et al.*, 2010)

Nooj : analyse de corpus utilisant des ressources sous la forme de dictionnaires et de grammaires

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS



Pourcentage de documents possédant un texte nettoyé

Constat 75% des documents possèdent un **texte nettoyé**

Outils CorText : outil d'analyse et d'exploitation de corpus textuels

TXM : outil d'analyse textométrique de corpus de textes (Heiden *et al.*, 2010)

Nooj : analyse de corpus utilisant des ressources sous la forme de dictionnaires et de grammaires

Requête `qualityIndicators.tdmReady:true`

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS

NOMBRE DE MOTS DU PDF ^

Nombre de mots (0 à 102 111)

à

NOMBRE DE CARACTÈRES DU PDF ^

Nombre de caractères (0 à 673 804)

à

Filtres Istex Search

Constats 2 documents de plus de 100 pages, 1 document de plus de 100 000 mots

Constitution d'un corpus spécialisé

EXAMEN DES INDICATEURS

NOMBRE DE MOTS DU PDF ^

Nombre de mots (0 à 102 111)

à

NOMBRE DE CARACTÈRES DU PDF ^

Nombre de caractères (0 à 673 804)

à

Filtres Istex Search

Constats 2 documents de plus de 100 pages, 1 document de plus de 100 000 mots

Requête `AND qualityIndicators.pdfPageCount:[* TO 100] AND qualityIndicators.pdfWordCount:[* TO 100000]`

Constitution d'un corpus spécialisé

CONCLUSION

Une méthodologie itérative : examen des données ↔ révision de la requête

- La pertinence scientifique : circonscrire au maximum le sujet quitte à créer du silence ?

Constitution d'un corpus spécialisé

CONCLUSION

Une méthodologie itérative : examen des données ↔ révision de la requête

- La pertinence scientifique : circonscrire au maximum le sujet quitte à créer du silence ?
- La pertinence technique : se limiter aux documents possédant un résumé, un PDF texte, un texte nettoyé, un nombre de mots / pages raisonnable ?

Constitution d'un corpus spécialisé

CONCLUSION

Une méthodologie itérative : examen des données ↔ révision de la requête

- La pertinence scientifique : circonscrire au maximum le sujet quitte à créer du silence ?
- La pertinence technique : se limiter aux documents possédant un résumé, un PDF texte, un texte nettoyé, un nombre de mots / pages raisonnable ?

Des aides précieuses développées dans Istex

- L'interface Istex Search

Constitution d'un corpus spécialisé

CONCLUSION

Une méthodologie itérative : examen des données ↔ révision de la requête

- La pertinence scientifique : circonscrire au maximum le sujet quitte à créer du silence ?
- La pertinence technique : se limiter aux documents possédant un résumé, un PDF texte, un texte nettoyé, un nombre de mots / pages raisonnable ?

Des aides précieuses développées dans Istex

- L'interface Istex Search
- L'appui de Lodex et des web services

Constitution d'un corpus spécialisé

CONCLUSION

Une méthodologie itérative : examen des données ↔ révision de la requête

- La pertinence scientifique : circonscrire au maximum le sujet quitte à créer du silence ?
- La pertinence technique : se limiter aux documents possédant un résumé, un PDF texte, un texte nettoyé, un nombre de mots / pages raisonnable ?

Des aides précieuses développées dans Istex

- L'interface Istex Search
- L'appui de Lodex et des web services
- Le catalogue [TM Tools explorer](#) (liste des outils, de leurs fonctionnalités et contraintes)



3. Conclusion

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM
- **À venir** : 138 ebooks (2011-2023) en poésie, histoire de l'art, architecture ou linguistique (Torrossa Edición, Casalini)

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM
- **À venir** : 138 ebooks (2011-2023) en poésie, histoire de l'art, architecture ou linguistique (Torrossa Edición, Casalini)

Un écosystème à votre service

- Des nouveaux web services sur Istex TDM (e.g. [entityTag](#))

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM
- **À venir** : 138 ebooks (2011-2023) en poésie, histoire de l'art, architecture ou linguistique (Torrossa Edición, Casalini)

Un écosystème à votre service

- Des nouveaux web services sur Istex TDM (e.g. [entityTag](#))
- Des nouvelles terminologies sur Loterre (e.g. [vocabulaire de TAL](#))

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM
- **À venir** : 138 ebooks (2011-2023) en poésie, histoire de l'art, architecture ou linguistique (Torrossa Edición, Casalini)

Un écosystème à votre service

- Des nouveaux web services sur Istex TDM (e.g. [entityTag](#))
- Des nouvelles terminologies sur Loterre (e.g. [vocabulaire de TAL](#))
- Nouvelle application Istex Search

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM
- **À venir** : 138 ebooks (2011-2023) en poésie, histoire de l'art, architecture ou linguistique (Torrossa Edición, Casalini)

Un écosystème à votre service

- Des nouveaux web services sur Istex TDM (e.g. [entityTag](#))
- Des nouvelles terminologies sur Loterre (e.g. [vocabulaire de TAL](#))
- Nouvelle application Istex Search
- Nouvelle version de Lodex

Conclusion

Istex

- Réservoir de plus de 28M de publications multilingues et multidisciplinaires
- Données apprêtées pour le TDM
- **À venir** : 138 ebooks (2011-2023) en poésie, histoire de l'art, architecture ou linguistique (Torrossa Edición, Casalini)

Un écosystème à votre service

- Des nouveaux web services sur Istex TDM (e.g. [entityTag](#))
- Des nouvelles terminologies sur Loterre (e.g. [vocabulaire de TAL](#))
- Nouvelle application Istex Search
- Nouvelle version de Lodex
- **À venir** : enrichissement de l'application Istex Search, création d'une plateforme dédiée aux corpus Istex



4. Liens, contacts, références

Adresses utiles

Se connecter

- Istex : <http://www.istex.fr>
- Istex Search : <https://search.istex.fr>
- Istex Data : <https://data.istex.fr>
- Lodex : <https://www.lodex.fr>
- Istex TDM : <https://services.istex.fr>

S'authentifier

- Vérifier ses droits d'accès : <https://api.istex.fr/auth>

Documentation & tutoriels

Se documenter

- Documentation Usage TDM d'Istex : <https://doc.istex.fr/tdm>
- Documentation API Istex : <https://doc.istex.fr/api>
- Documentation Lodex : <https://www.lodex.fr/docs/documentation>

Se former

- Tutos Lodex : <https://callisto-formation.fr/course/view.php?id=194>

Informations & contacts

Se tenir informé

- Article d'actualité : <https://www.istex.fr/category/actualites>
- Plateforme X : @ISTEX_Platform

Chercher de l'aide / Contribuer à l'amélioration

- Via le formulaire : <https://www.istex.fr/contact>
- Via la liste : contact@listes.istex.fr

Références

- ATANASSOVA, I. et BERTIN, M. (2023). Multilinguism in References: a study of the ISTEK Dataset. *In Proceedings of ISSI 2023 – the 19th International Conference of the International Society for Scientometrics and Informetrics*, volume 3, pages 9–10, Bloomington, Indiana, US.
- BANNOUR, N., SERVAN, C., NÉVÉOL, A. et TANNIER, X. (2024). A Benchmark Evaluation of Clinical Named Entity Recognition in French.
- BOUCHARD, A. (2024). Sortir de Google Scholar, Scopus ou Web of Science : que valent Lens, Dimensions, OpenAlex et Matilda pour la recherche bibliographique ? [à distance] | URFIST de Paris.
- BÉRARD, R. (2012). ISTEK. Vers des services innovants d'accès à la connaissance. *Arabesques*, (66):6–8.
- CUXAC, P. (2024). La fouille de textes en IST : les outils Istex-TDM. Nancy, France.
- CUXAC, P., COLLIGNON, A., GREGORIO, S. et PARMENTIER, F. (2019). Des bases de données massives au Web de données : désambiguïsation et alignement d'entités géographiques dans les textes scientifiques. *In 12ème Colloque international d'ISKO-France : Données et mégadonnées ouvertes en SHS : de nouveaux enjeux pour l'état et l'organisation des connaissances ?*, Montpellier, France.

Références

- de SALABERT, C. et BARREAUX, S. (2020). Vers un corpus optimal pour la fouille de textes : stratégie de constitution de corpus spécialisés à partir d'ISTEX. In BENZITOUN, C., BRAUD, C., HUBER, L., LANGLOIS, D., OUNI, S., POGODALLA, S. et SCHNEIDER, S., éditeurs : *Actes de la 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 4 : Démonstrations et résumés d'articles internationaux*, pages 66–69, Nancy, France. ATALA et AFCP.
- GREGORIO, S., COLLIGNON, A., PARMENTIER, F. et THOUVENIN, N. (2019). LODEX : des données structurées au web sémantique. In *Atelier Web des Données de la 19ème Conférence sur l'Extraction et la Gestion des Connaissances (EGC 2019)*, Metz, France.
- GUSENBAUER, M. (2022). Search where you will find most: Comparing the disciplinary coverage of 56 bibliographic databases. *Scientometrics*, 127(5):2683–2745.

Références

- HEIDEN, S., MAGUÉ, J.-P. et PINCEMIN, B. (2010). TXM : Une plateforme logicielle open-source pour la textométrie - conception et développement. *In Statistical Analysis of Textual Data - Proceedings of 10th International Conference Journées d'Analyse statistique des Données Textuelles*, volume 2, pages 1021–1032, Rome, Italie. Edizioni Universitarie di Lettere Economia Diritto. Issue: 3.
- KHAYARI, M., RESZETKO, V., VACHEZ, D., VEDOVOTTO, N., YON, J. et AUBIN, S. (2021). De TermSciences à Loterre : comment l'Inist-CNRS a rendu les terminologies ouvertes plus conformes aux principes FAIR.
- NIEDERLENDER, C. (2017). L'API Istex : le sésame pour accéder aux ressources acquises. *Arabesques*, (84):17–19.
- PETITJEAN-MONNIN, A., PIERROT, R. et FABRY, C. (2022). ISTEEX : de la plateforme de référence à l'infrastructure de recherche. *Arabesques*, (107):20–20. Publisher: Agence bibliographique de l'enseignement supérieur.
- RAUF, S. A. et YVON, F. (2024). Translating scientific abstracts in the bio-medical domain with structure-aware models. *Computer Speech & Language*, 87:1–19.