



Constitution et analyse d'un corpus documentaire à partir d'Istex

Écoles doctorales de l'Université de Lorraine – 20 janvier 2026
mathilde.huguin@inist.fr - emeline.caule@inist.fr

Inist-CNRS

 Inist – 2025



Programme & Objectifs

Recherche documentaire & Text and data mining

Mardis 13 janvier et 20 janvier de 9h30 à 12h30

1. Connaître le réservoir multidisciplinaire Istex
2. Savoir interroger le réservoir
3. Acquérir les bases pour manipuler l'outil de datavisualisation Lodex & des web services
4. Prendre du recul sur la constitution de corpus et la fouille de textes



Documents utiles

Télécharger ce support et les fichiers utiles aux manipulations sur la page

<https://universite-lorraine.formation.lodex.fr/instance/supports>

Programme



01

Introduction : le TDM & Istex

02

Constitution d'un corpus
spécialisé

03

Manipulations des outils
présentés

A grayscale photograph of a stack of books. The top book is open, showing its pages. Below it are several closed books. In the background, a laptop is visible, slightly out of focus. The overall scene suggests a study or research environment.

1. Introduction : le TDM & Istex

Introduction : le TDM & Istex

« Toute technique d'analyse automatisée visant à analyser des textes et des données sous une forme numérique afin d'en dégager des informations, ce qui comprend, à titre non exhaustif, des constantes, des tendances et des corrélations »

(Ordonnance du 24 novembre 2021, [ouvrirlascience.fr](https://www.ouvrirlascience.fr))

The logo consists of a green oval containing the text "MaTOS" in blue.

MaTOS

- Extraire de l'information pertinente (recherche d'information)
- Analyser de gros volumes de textes
- Détecter des sentiments dans les textes
- Construire des résumés automatiques
- Désambiguïser des lieux, des personnes...
- Faire des systèmes de recommandations
- Détecter des « fake news »
- Trier des mails, des textes
- Faire de la bibliométrie

Introduction : le TDM & Istex

Évolution massive de la quantité de publications

- Sur ces six dernières années, la part des publications scientifiques mondiales a augmenté de 47% (Hanson *et al.*, 2023)
- Selon les domaines entre 12% (médecine) et 82% (sciences humaines) des articles ne sont jamais cités ([Larivière & Gingras, 2009](#))
- 50% des articles cités ne sont jamais lus ([Office of Research Integrity](#))

Évolution massive des techniques de TAL et d'IA

- Accompagnée par l'implication d'industriels (intérêt financier) : ex. chatGPT (OpenAI)
- Évolution de la puissance de calcul et de stockage (super ordinateur Jean Zay : 125,9 pétaflops, soit 125,9 millions de milliards d'opérations par seconde)

Introduction : le TDM & Istex

Évolution de la législation

- 2019 : directive européenne sur le droit d'auteur > exceptions au code de la propriété intellectuelle dans un cadre de recherche
- 2021-2022 : transposition de la directive européenne en droit français

Évolution des moyens pour la science ouverte

- 2018 : 1^{er} plan pour la science ouverte 5M/an
- 2021 : 2^{ème} plan national pour la science ouverte 15M/an

Introduction : le TDM & Istex

Des données

Des spécialistes

Des outils de visualisation

Des outils de traitement



Introduction : le TDM & Istex

Exemple de traitements : [le WS catégories scientifiques Inist](#)

Classification supervisée : apprentissage sur des données déjà labellisées

- Input : l'Inist dispose d'un fond de documents labellisés suivant leur domaine scientifique
- Action : l'algorithme apprend seul à associer un document à un label
- Output : le programme associe à chacun des nouveaux documents présenté un label

Introduction : le TDM & Istex

```
[
  {
    "idt": "08-040289",
    "value": "Planck 2015 results. XIII. Cosmological parameters.We present"
  }
]
```

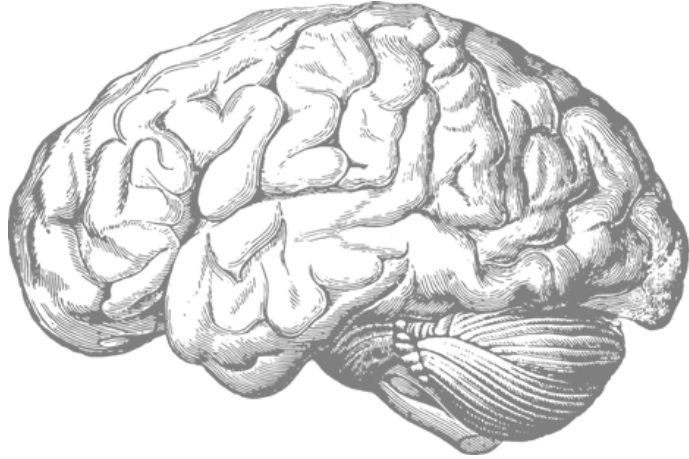
```
[{
  "idt": "08-040289",
  "value": [
    {
      "code": {
        "id": "001",
        "value": "Sciences exactes et technologie."
      },
      "confidence": 1.0000057220458984,
      "rang": 1
    },
    {
      "code": {
        "id": "001E",
        "value": "Terre, océan, espace."
      },
      "confidence": 0.9999549388885498,
      "rang": 2
    },
    {
      "code": {
        "id": "001E03",
        "value": "Astronomie."
      },
      "confidence": 1.0000100135803223,
      "rang": 3
    }
  ]
}]
```

A grayscale photograph of a stack of books. The top book is open, showing its pages. Below it are several closed books. In the background, a laptop is visible, slightly out of focus. The overall scene suggests a library or a study environment.

2. Constitution d'un corpus spécialisé

Cas d'usage | Requêtage
Téléchargement | Exploration

Cas d'usage



Besoin : *recherche en collaboration avec des chercheurs en psychologie et en neurosciences pour constituer un thésaurus sur la mémoire humaine pour : indexer des publications ou des ressources ; uniformiser la terminologie du domaine ; répertorier les concepts issus de la recherche cognitive sur la mémoire humaine*

Méthodologie

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

Méthodologie



**2 phases
&
2 outils**

1

**Pertinence
scientifique**

Istex Search : construction
et affinement requête,
extraction corpus v0

Lodex : examen de la
pertinence du corpus grâce aux
datavisualisations & aux
facettes

2

**Exploitabilité
TDM**

Istex Search : examen des
indicateurs techniques
pertinents

Requêtage

Requêtage

memory OR metamnesy OR metamemory OR
mémoire OR métamnésie OR métamémoire

Résultats (21-10-2024) : 1 701 374 doc.

Explication de la requête

- Les mots sont recherchés dans tout le document (métadonnées, texte intégral, références bibliographiques, enrichissements)
- Insensibilité à la casse
- La syntaxe Lucene autorise 3 opérateurs booléens (AND, OR, NOT) et par défaut l'espace blanc équivaut à un OR

Requêtage

m?mor* mnesic metamne* m?tam?mor*
mémoire* mnési* mnémo* métamnési*

Résultats (21-10-2024) : 2 664 987 doc.

Explication de la requête

- Rechercher des variantes morphosyntaxiques et des dérivés des mots-clés grâce à la troncature (*)
- Mutualiser certains termes (é|e) grâce au joker (?)

Requêtage

ENFANCES
FAMILLES
GÉNÉRATIONS

Enfances Familles Générations

Revue interdisciplinaire sur la famille contemporaine

The COVID Families

La pandémie de
La pandemia de

McMillan, I. F., L. M. Armstrong et J. Langhinrichsen-Rohling. 2021. «Transitioning to parenthood during the pandemic: covid-19 related stressors and first-time expectant mothers' mental health», *Couple and Family Psychology: Research and Practice*, 10(3): 179-189.

McRae, C. S., N. C. Overall, A. M. E. Henderson, ... et V. T. Chang. 2021. «Parents' distress and poor parenting during a covid-19 lockdown: the buffering effects of partner support and cooperative coparenting», *Developmental Psychology*, 57(10): 1623-1632.

Meloche, K. 2022. *Les impacts des mesures sanitaires en contexte de la COVID-19 sur les personnes âgées vivant dans les centres d'hébergement de soins de longue durée au Canada (mémoire)*, École de service social, Université d'Ottawa.

Laurence Charton, Léoni Labrecque and Joseph Josy Lévy

Requêtage

Afrique

Afriques

New Stud
Garima E

Nouvelles études
Garima

Sergey Kim

The study was financed by the project ANR EthioChrisProcess: Christianisation et interactions religieuses en Éthiopie (VIe-XIIIe s.), in 2019-2020, and IRIS Scripta-PSL. Histoire et pratiques de l'écrit, in 2020-2021 (<https://scripta.psl.eu/projets/lectures-liturgiques-et-usages-memoriels-des-manuscrits-des-evangiles-dabba-garima-ethiopie-vi-xvie-s/>). I warmly thank the project supervisors Anaïs Wion and Marie-Laure Derat for scientific advice and bibliographical help, as well as Alessandro Bausi for helping obtain reproductions of the Abba Garima Gospels and Nafisa Valieva for fruitful discussions. Special thanks go to Michael Gervers for giving permission to use his photographs of manuscripts and to Alexei Pentkovsky for consulting on liturgical matters. I am grateful to two anonymous reviewers who helped a lot to correct mistakes and elucidate the exposition.

Requêtage



Ren

Revue des mondes musulmans et de la

Romanciers, poètes, philosophes et historiens trouvent dans *al-Yawm al-sābi'* l'espace intellectuel de leur convergence, un lieu de médiation facilitateur de leurs échanges et de la circulation transnationale (à l'échelle du monde arabe) de leurs productions écrites. Nombreux sont ceux à y avoir tenu une rubrique pendant plusieurs mois ou années, l'une des plus mémorables étant celle du poète palestinien Mahmoud Darwich¹¹. Sa rubrique est d'ailleurs enrichie d'une correspondance littéraire qu'il entretient avec le poète palestinien Samih Al-Qasim, à partir du mois de mai 1986, et ce pendant deux années. Dans un registre lyrique, leurs échanges

Sociol
panar

Sociohistory of al-Yawm al-sābi': a pan-Arab magazine in exile (1984-1991)

التاريخ الاجتماعي لـ « اليوم السابع », مجلة عربية في المنفى (1991-1984)

Requêtage

title:(m?mor* mnesic metamne* m?tam?mor*
mémoire* mnési* mnémo* métamnési*)

Résultats (21-10-2024) : 85 689 doc.

Explication de la requête

- Cibler le champ interrogé
- Les noms de champs sont introduit par :

Requêtage



L'Atelier du Centre de recherches historiques

Revue électronique du CRH

Les thèses du CRH | 2022

Déposer les armes, prendre la parole. La lutte armée d'extrême gauche en Italie (1969-2015) : narrations autobiographiques, représentations historiques, controverses mémorielles

Thèse soutenue le 29 novembre 2022

Laying Down the Guns, Taking up Writing. Left-Wing Armed Struggle in Italy (1969 - 2015): Autobiographical Narratives, Historical Representations, Memorial Controversies

Lorenzo De Sabbata

Requêtage

```
title:(m?mor* mnesic metamne* m?tam?mor*  
mémoire* mnési* mnémo* métamnési*) AND  
categories.wos.raw:("2 - neurosciences" "2 -  
psychology" "2 - clinical neurology" "2 -  
psychology, biological" "2 - psychiatry" "2 -  
psychology, clinical" "2 - geriatrics &  
gerontology" "2 - gerontology" "2 -  
neuroimaging" "2 - behavioral sciences" "2 -  
psychology, experimental")
```

Résultats (21-10-2024) : 20 726 doc.

Explication de la requête

- Cibler les catégories scientifiques interrogées grâce au champ categories.wos.raw

Requêtage

ON MEMORY.¹

BY JAMES ROSS, M.D., LL.D.

THE subject for discussion to-night is "Memory," and I presume that we are supposed to enquire into its nature, its limitations from other modes of consciousness, its laws of action, its connection with cerebral activity, and its dissolution by disease. But memory is so intertwined with every other form of mental activity, with the feelings which accompany instinctive and acquired movements, with sensations, perceptions and conceptions, with reasoning, from the simplest to the most abstract, and even with the emotions and the will, that it is simply impossible to deal with it satisfactorily in the few moments at my disposal to-night. To treat this subject with even a moderate degree of adequacy would involve the writing of a complete system of psychology, and consequently in the few remarks I have to offer I will

My *first* proposition with regard to memory is that it is a mode of consciousness, and consequently to speak of unconscious memory is a contradiction in terms. If a psychical act or an impression on the organism is wholly unperceived it cannot afterwards be recalled as a memory.

My *second* proposition is that our primary knowledge of memory is obtained by means of introspection. This knowledge is no doubt greatly extended by objective observation of the actions of other organisms, but my guarantee that anyone except myself possesses memory

¹ Read at a meeting of the Neurological Society in December, 1890.

Requêtage

```
title:(m?mor* mnesic metamne* m?tam?mor*  
mémoire* mnési* mnémo* métamnési*) AND  
categories.wos.raw:("2 - neurosciences" "2 -  
psychology" "2 - clinical neurology" "2 -  
psychology, biological" "2 - psychiatry" "2 -  
psychology, clinical" "2 - geriatrics &  
gerontology" "2 - gerontology" "2 -  
neuroimaging" "2 - behavioral sciences" "2 -  
psychology, experimental") AND  
publicationDate:[2000 TO *]
```

Résultats (07-01-2026) : 20 290 doc.

Explication de la requête

- Limiter la recherche aux 20 dernières années grâce au champ `publicationDate`
- Utiliser un intervalle `[2000 TO *]`



Téléchargement

Téléchargement

```
title:(m?mor* mnesic metamne* m?tam?mor*  
mémoire* mnési* mnémo* métamnési*) AND  
categories.wos.raw:("2 - neurosciences" "2 -  
psychology" "2 - clinical neurology" "2 -  
psychology, biological" "2 - psychiatry" "2 -  
psychology, clinical" "2 - geriatrics &  
gerontology" "2 - gerontology" "2 -  
neuroimaging" "2 - behavioral sciences" "2 -  
psychology, experimental") AND  
publicationDate:[2000 TO *]
```


Téléchargement

Configurez votre téléchargement

USAGE PERSONNALISÉLODEXCORTEXTGARGANTEXT

☐ Texte intégral

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

☒ Métadonnées

- ☒ JSON
- ☐ XML
- ☐ MODS
- ☐ Annexes
- ☐ Couvertures

☐ Enrichissements

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

Trier par : pertinence & qualité ▼

Télécharger 13 158 / 13 158

Format de l'archive : ZIP ▼Compression : moyenne ▼

TÉLÉCHARGER

Lodex

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

[EN SAVOIR PLUS](#)

Requête

```
title:(m?mor* mnesic metamne* m?tam?mor* mémoire* mnési* mnémo* métamnési*) AND categories.wos.raw:("2 - neurosciences" "2 - psychology" "2 - clinical neurology" "2 - psychology, biological" "2 - psychiatry" "2 - psychology, clinical" "2 - geriatrics &...
```

Requête brute complète

```
https://api.istex.fr/document?q=title:(m?mor* mnesic metamne* m?tam?mor* mémoire* mnési* mnémo* métamnési*) AND categories.wos.raw:("2 - neurosciences" "2...
```

Exploration du corpus V0

■ New Visitor ■ Returning Visitor



2 phases
&
2 outils

1

Pertinence
scientifique

Istex Search : construction
et affinement requête,
extraction corpus v0

Lodex : examen de la
pertinence du corpus grâce aux
datavisualisations & aux
facettes

2

Exploitabilité
TDM

Istex Search : examen des
indicateurs techniques
pertinents

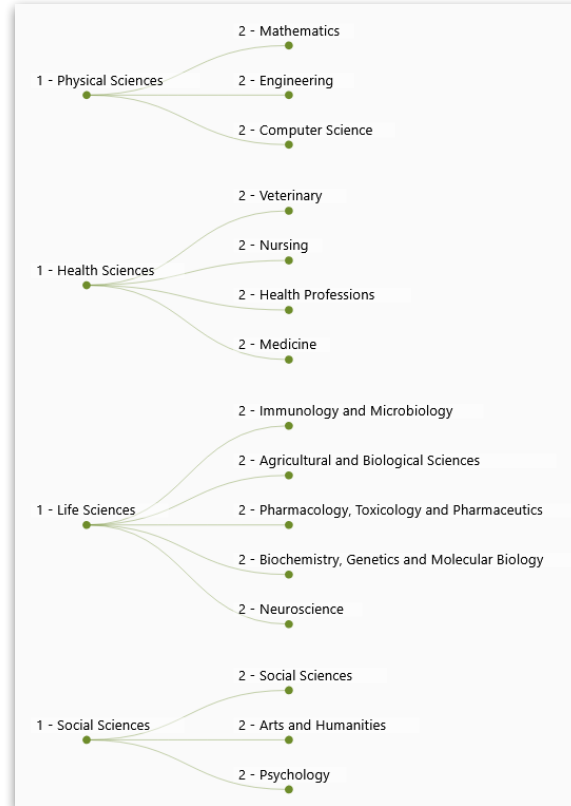
Exploration du corpus V0

Années de publication



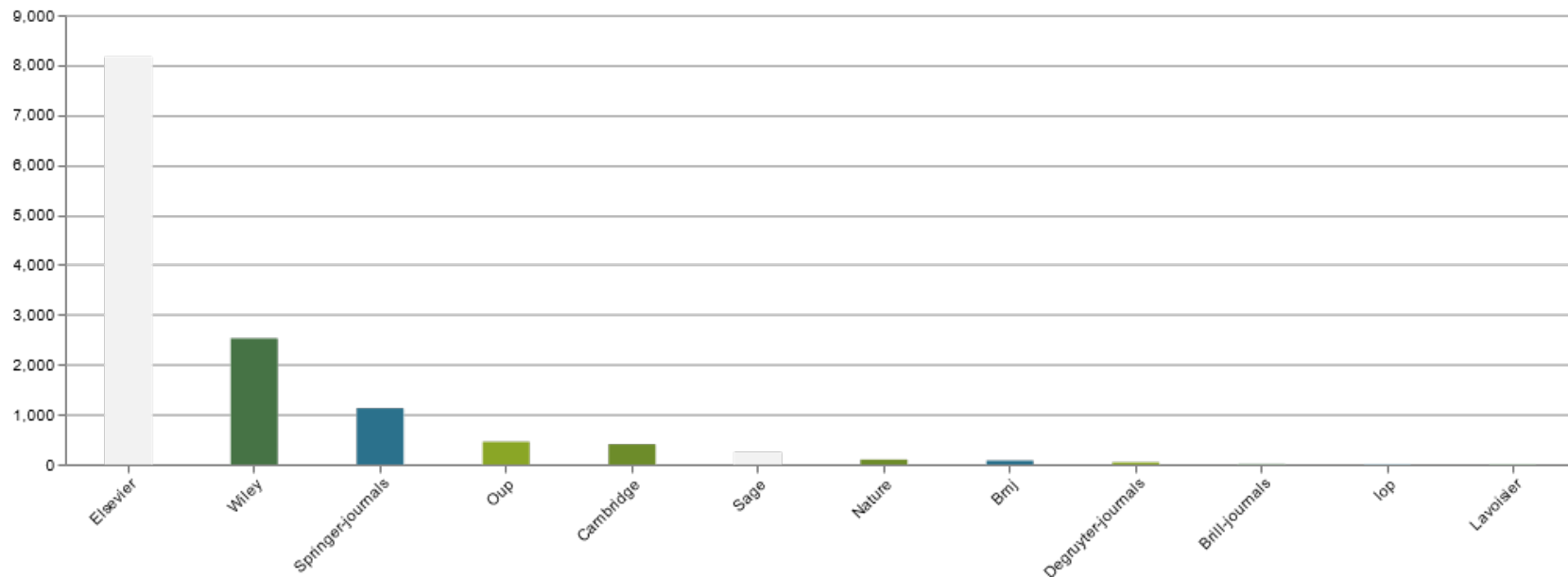
Exploration du corpus V0

Catégories scientifiques
Scopus



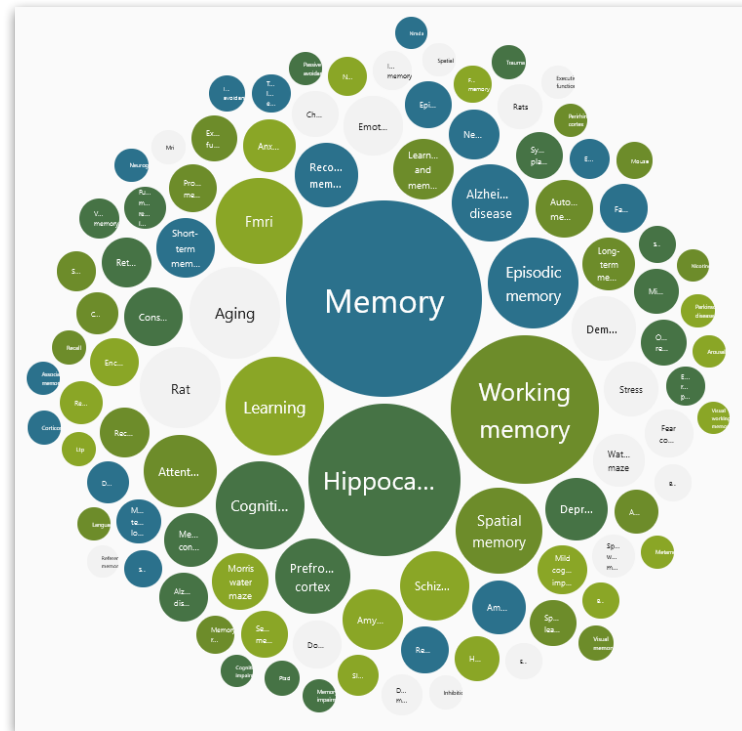
Exploration du corpus V0

Éditeurs



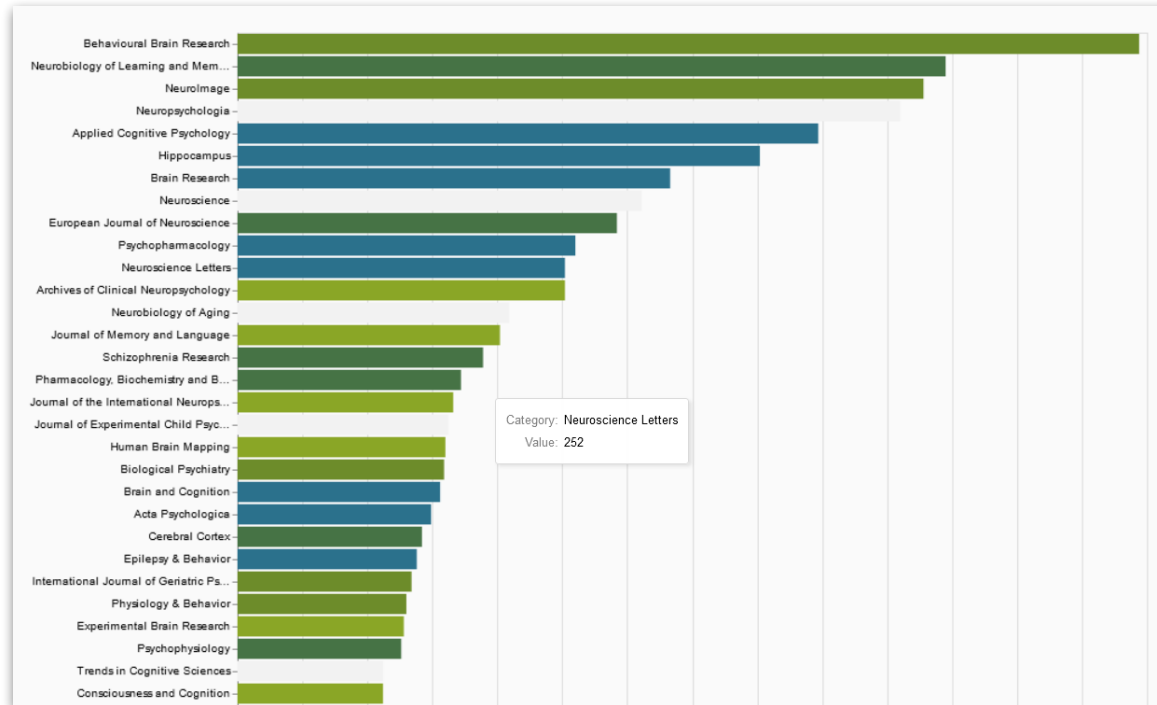
Exploration du corpus V0

Mots-clés d'auteur



Exploration du corpus V0

Revues (346)



Exploration du corpus V0

Revue d'éthologie

Revue / Monographie (3) ^

☐ exclure

animal|

VALEUR COMPTER ↓

TOUT COCHER

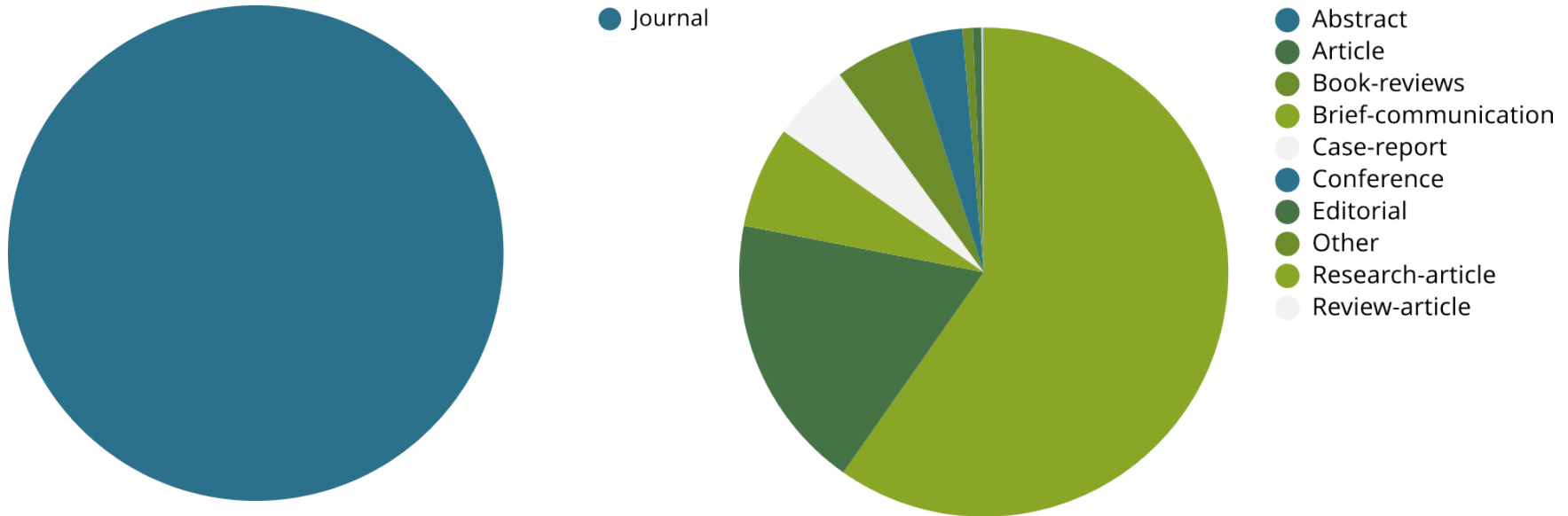
<input type="checkbox"/> Animal Cognition	51
<input type="checkbox"/> Animal Behaviour	43
<input type="checkbox"/> Applied Animal Behaviour Science	7

Showing 1 to 3 of 3 < >



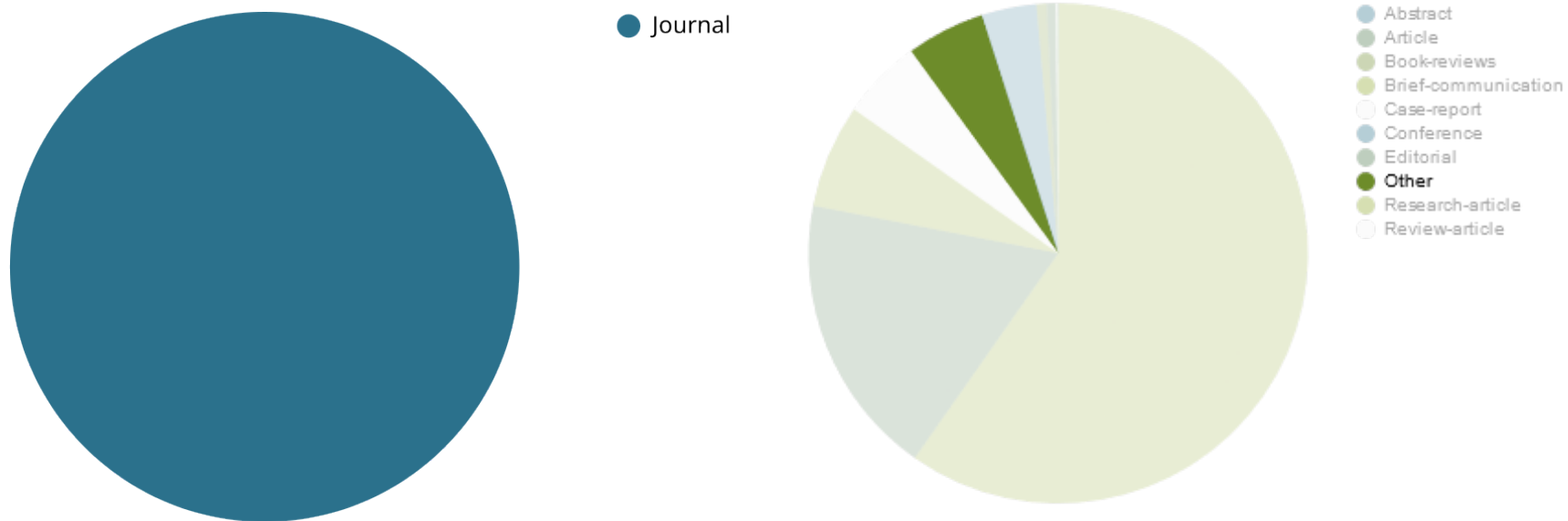
Exploration du corpus V0

Types de publication et de document



Exploration du corpus V0

Types de publication et de document



Exploration du corpus V0

Types de publication et de document

Journal of the International Neuropsychological Society (2000), 6, 376–377.
Copyright © 2000 INS. Published by Cambridge University Press. Printed in the USA.

In Memoriam

Laird S. Cermak, Ph.D.

June 24, 1942–November 4, 1999

Laird Cermak was to become President of INS in February 2000, taking over the responsibilities for leading a scientific society he loved and served for over a quarter of a century. In many ways he embodied the very values the INS has always hoped to promote in excellence of scholarship, keen interest in behavioral neurosciences, and an easy and reflexive interdisciplinary citizenship. His name has been synonymous with a distinctive line of intellectual inquiry in amnesia and memory mechanisms for as long as I have been in neuropsychology myself.

It now can be told that I was alerted to Laird's considerable intellectual talent even as a graduate student. I recall that my major Professor, Phillip Rennick, showed me the initial application for research support he was reviewing that Laird had made to the Veterans Administration in the early 1970s. "Read this really excellent grant, Ken," he said, "because this guy is really going to do some significant work." Given that Phil tended to view proposals as having fatal flaws usually no later than the literature review, I was impressed and learned with awe what moved Phil. Apparently many others have joined in Phil's appraisal and appreciation of a career of scholarly promise fulfilled. I know that I did from that day. Laird took a central faculty role in one of the most protean research groups in the history of neuropsychology.



À la mémoire de Betty Havens 9 octobre 1936 au 1^{er} mars 2005

Marcia Finlayson,¹ Neena Chappell,² and Lynn McDonald³

- 1 Université de l'Illinois à Chicago
- 2 Université de Victoria
- 3 Université de Toronto



lège qui nous a été offert de partager une partie de son

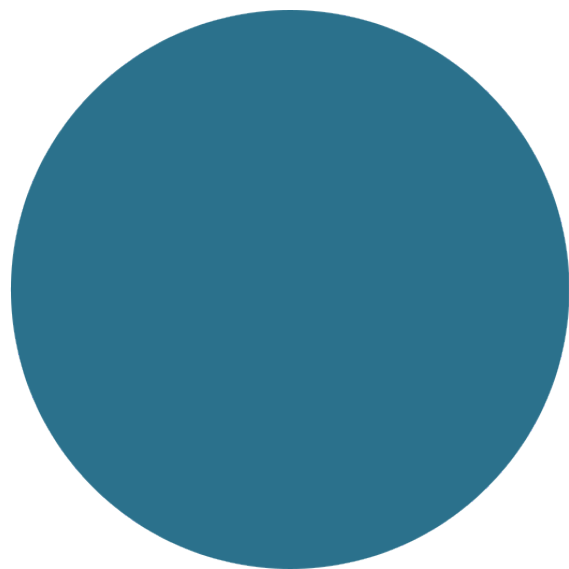
Lorsque Betty Havens est décédée, le 1^{er} mars au matin, la nouvelle s'est propagée rapidement que notre chère amie, collègue et mentor, celle qui nous avait toujours soutenus, était partie. Nombreux sont ceux à qui manqueront son rire chaleureux, ses répliques amusantes et ses petits clins d'œil. Nous sommes reconnaissants du privi-

Nous connaissons son engagement et son empressement constant à lutter pour les aînés et la gérontologie. Son intérêt envers la gérontologie était indiscutable, même à la toute fin, puisqu'elle lisait *The Gerontologist* dans son lit aux soins palliatifs.

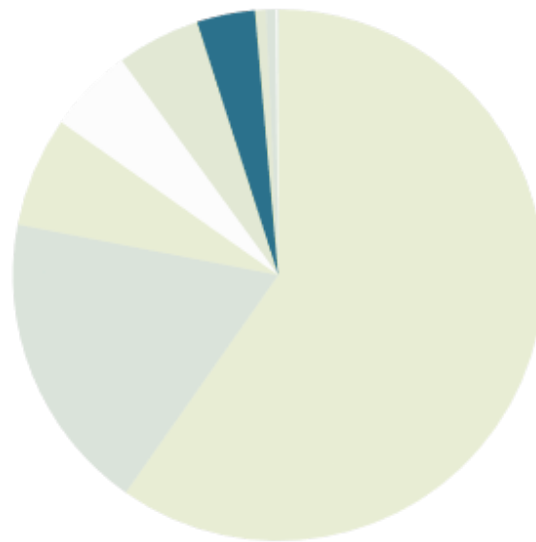
J'aimerais parler d'elle d'un point de vue plus personnel. Bien qu'elle n'était pas très démonstrative, elle soutenait sans aucune réserve les personnes en lesquelles elle croyait. Je me souviens très clairement de moments, pendant les premières années de ma carrière, où elle est venue me prendre par la main, lors de réunions du GSA, de l'ISA ou d'autres occasions, pour me présenter à Bernice Neugarten, à Matilda White Riley et à de nombreux autres grands spécialistes de la gérontologie : « Nous avons finalement l'un des grands de Montréal », disait-elle. Elle nous a

Exploration du corpus V0

Types de publication et de document



● Journal



- Abstract
- Article
- Book-reviews
- Brief-communication
- Case-report
- Conference
- Editorial
- Other
- Research-article
- Review-article

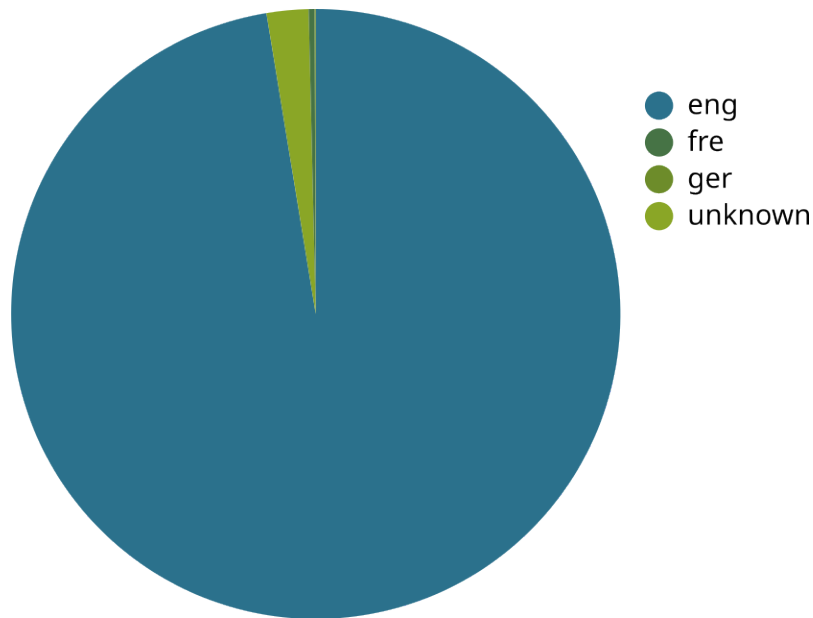
Types de publication et de document

1498

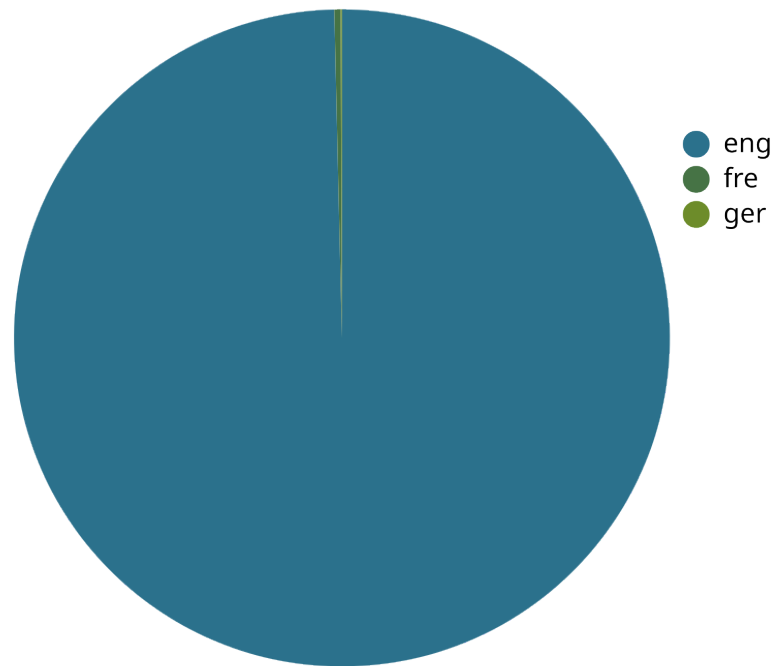
42

Exploration du corpus V0

Langues



Langues dans Istex (292 unknown)



Résultat du web service

Exploration du corpus V0

Bilan



01

Revue

NOT host.title:("Animal Cognition"
"Animal Behaviour" "Applied Animal
Behaviour Science")



02

**Types de
documents**

NOT title:("in mémoires" "à la
mémoire") NOT genre:abstract



03

Langues

NOT language:ger

Révision de la requête

Suppression de 578 documents

```
title:(m?mor* mnesic metamne* m?tam?mor*
mémoire* mnési* mnémo* métamnési*) AND
categories.wos.raw:("2 - neurosciences" "2 -
psychology" "2 - clinical neurology" "2 -
psychology, biological" "2 - psychiatry" "2 -
psychology, clinical" "2 - geriatrics &
gerontology" "2 - gerontology" "2 -
neuroimaging" "2 - behavioral sciences" "2 -
psychology, experimental") AND
publicationDate:[2000 TO *]
```

```
title:(m?mor* mnesic metamne* m?tam?mor* mémoire*
mnési* mnémo* métamnési*) AND categories.wos.raw:("2
- neurosciences" "2 - psychology" "2 - clinical neurology" "2 -
psychology, biological" "2 - psychiatry" "2 - psychology,
clinical" "2 - geriatrics & gerontology" "2 - gerontology" "2 -
neuroimaging" "2 - behavioral sciences" "2 - psychology,
experimental") AND publicationDate:[2000 TO *] NOT
host.title:("Animal Cognition" "Animal Behaviour"
"Applied Animal Behaviour Science") NOT title:("in
mémoriam" "à la mémoire") NOT genre:abstract NOT
language:ger
```

Exploitabilité TDM

**2 phases
&
2 outils**

1

**Pertinence
scientifique**

Istex Search : construction
et affinement requête,
extraction corpus v0

Lodex : examen de la
pertinence du corpus grâce aux
datavisualisations & aux
facettes

2

**Exploitabilité
TDM**

Istex Search : examen des
indicateurs techniques
pertinents

Exploitableté TDM

Résumé



Constat

- 6 % des documents n'ont [pas de résumé](#)

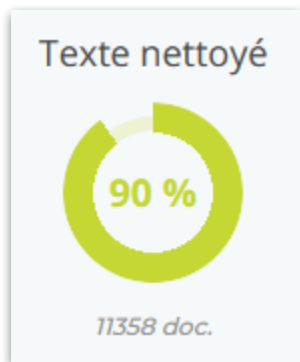
Procédures / outils

- Teeft : indexation non supervisée pour extraire les termes les plus spécifiques d'un contenu scientifique
- GarganText : plateforme collaborative d'exploration graphique de corpus

Requête AND abstract:*

Exploitabilité TDM

Texte nettoyé



Constat

- 10 % des documents n'ont pas de texte nettoyé ([texte brut](#) vs [texte nettoyé](#))

Procédures / outils

- CorText : outil d'analyse et d'exploitation de corpus textuels (script Terms extraction)
- TermSuite : outil d'extraction terminologique et d'alignement de termes multilingues

Requête AND `qualityIndicators.tdmReady:true`

Exploitabilité TDM

PDF texte



Constat

- Tous les documents ont un PDF texte (vs type [PDF image](#))

Procédures / outils

- Transformer le PDF en texte avant l'intégration dans l'outil souhaité ([tesseract](#))

Requête AND qualityIndicators.pdfText:true

Exploitabilité TDM

Nombre de pages / mots

Constats

- 25 documents possèdent plus de 100 000 mots
- 11 documents possèdent plus de 100 pages

Requêtes

AND qualityIndicators.pdfPageCount:[* TO 100]

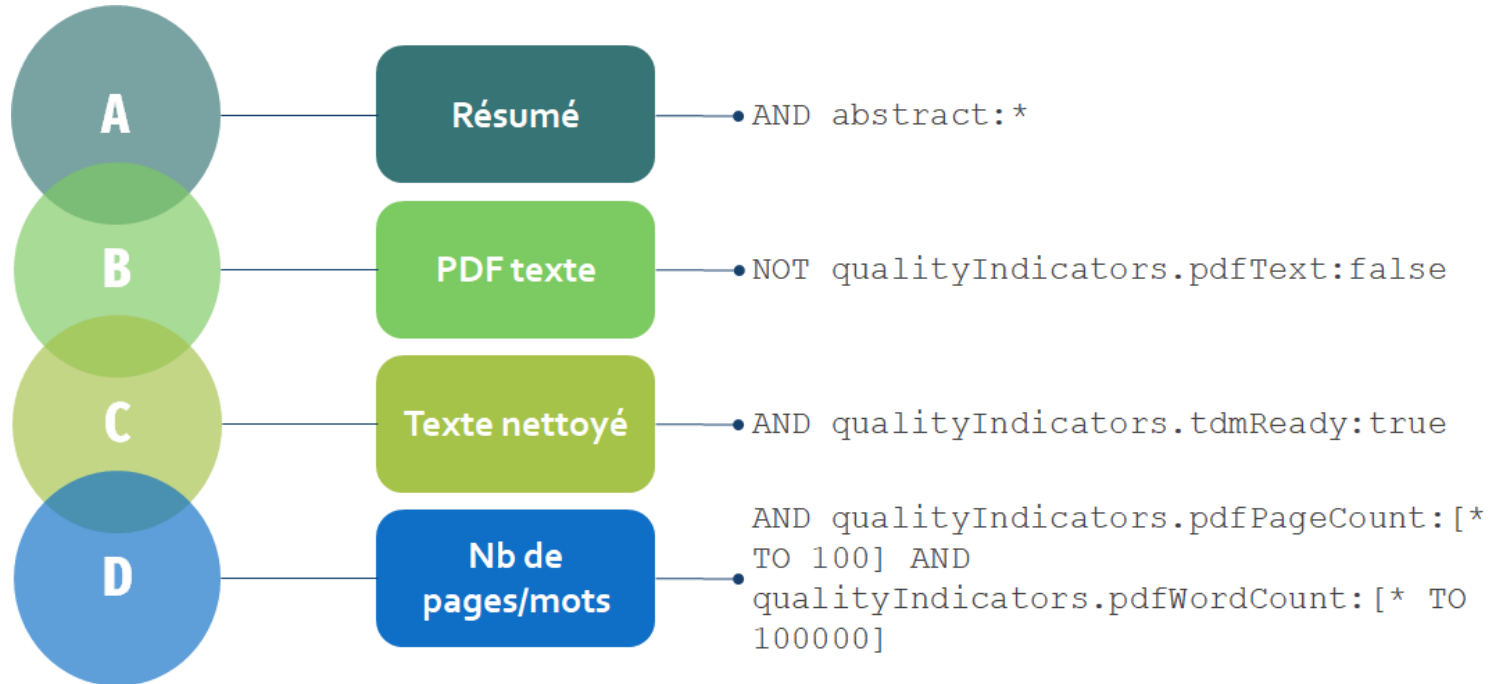
AND qualityIndicators.pdfWordCount:[* TO 100000]



The screenshot shows a modal dialog box with a light gray header containing the title "NOMBRE DE MOTS DU PDF" and a small upward-pointing arrow icon. Below the header, the text "Nombre de mots (106 179 à 161 114)" is displayed. Underneath, there are two input fields: the first contains "106 179" and the second contains "161 114", with a small "à" symbol between them. At the bottom of the dialog, there is a wide button labeled "APPLIQUER" and a smaller button with a close icon (an 'X' inside a square).

Exploitable TDM

Bilan




A black and white photograph showing a stack of several books. The top book is open, with its pages fanned out. To the right of the books, a portion of a laptop is visible, including its screen and keyboard area. The background is a solid, dark grey.

À vous !

Exercice 1 : construire une requête Istex

☐ **Étape 1** : Se rendre sur [Istex Search](#).

☐ **Étape 2** : Rechercher les formes anglaises et françaises *refugee*, *réfugié*, *asylum seeker* et *demandeur d'asile*.

Pour une aide sur la syntaxe des requêtes consulter les [Astuces de recherche](#)  (à droite dans Istex Search).

☐ **Étape 3** : Limiter le bruit et le silence.

☐ **Étape 4** : Répondre aux contraintes scientifiques et techniques.

Quelle est la taille de votre corpus ?

Exercice 1 : construire une requête Istex

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

Exercice 2 : chargement des données et création du site Lodex

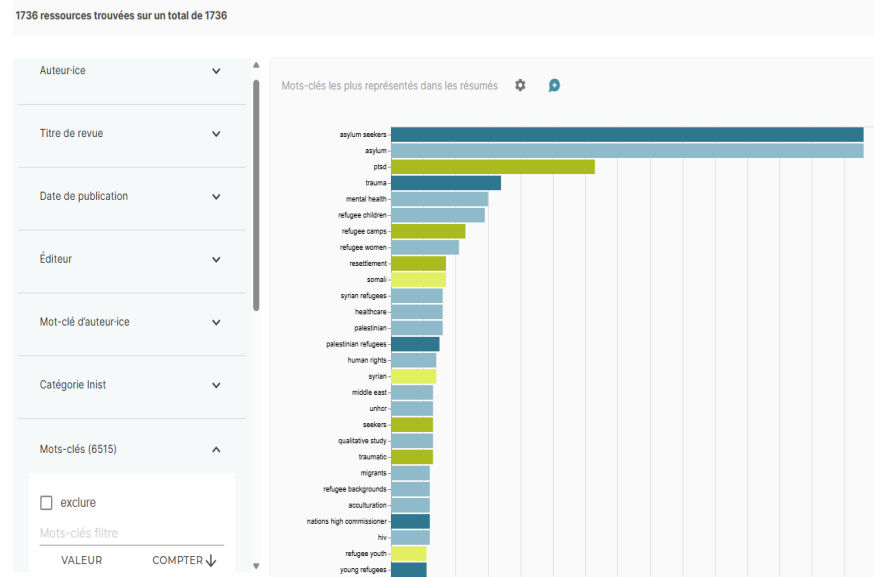
- ☐ Étape 1 : Télécharger le corpus.
- ☐ Étape 2 : Importer le corpus dans Lodex.
- ☐ Étape 3 : Importer le modèle.
- ☐ Étape 4 : Modifier un champ existant dans Lodex.

The screenshot displays the ISTEX website interface. At the top, there is a navigation bar with links: « ACCÈS ISTEX.FR », TOU WEB SERVICES, LOTERIE, CORPUS SCIENTIFIQUES, and ACTUALITÉS ISTEX. Below this, a header section features the text 'ISTEX exemple : Services' and 'exemple : Les technologies et les outils ISTEX pour les projets de recherche.' The main content area is titled 'Corpus Réfugié-es' and includes a sub-section 'Étude des réfugié-es politiques et climatiques'. This section contains an illustration of three globes and a text block stating: 'Le corpus Réfugié-es sert de base à une recherche en sociologie. Son objectif est de fournir un matériau de recherche pour analyser les **mouvements des réfugié-es à travers le monde** à travers la littérature scientifique. Il offre de distinguer la part de documents traitant des réfugié-es climatiques et des réfugié-es politiques et d'identifier leur origine géographique.' At the bottom, there are three informational boxes: 'Date de création' with a 'à renseigner' button, 'Structure' with the 'Université de Paris Cité' label, and 'Nombre de documents' showing the value '1736'.

Exercice 3 : Premiers pas vers le TDM

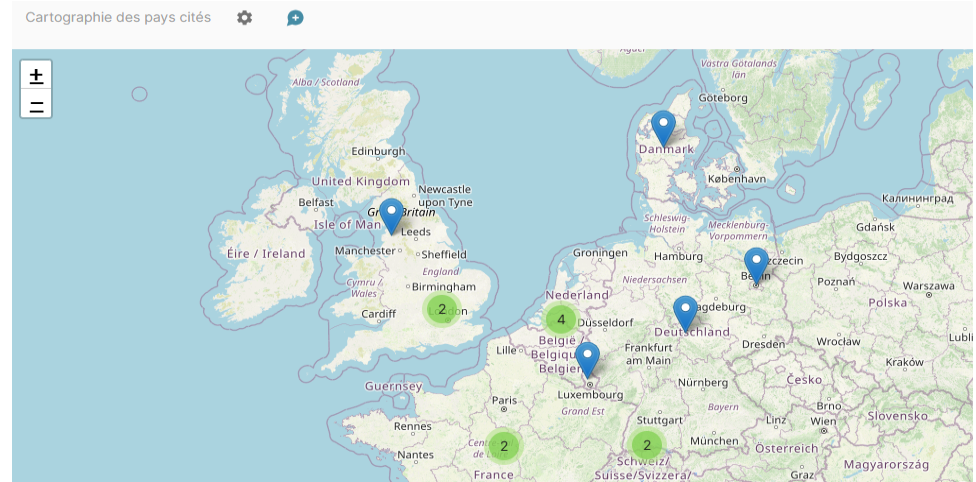
- ☐ **Étape 1** : Extraire les mots-clés des résumés via le web service Teeft.
- ☐ **Étape 2** : Créer un diagramme en barres.
- ☐ **Étape 3** : Créer une nouvelle facette Mots-clés.

Combien de documents comportent le terme *asylum seekers* et *environmental refugees* ?



Exercice 4 : Cartographie

- ☐ **Étape 1** : Récupérer les noms de lieux mentionnés.
- ☐ **Étape 2** : Uniformiser les noms de lieux pour créer une cartographie.
- ☐ **Étape 3** : Créer une carte en fonction des pays mentionnés dans les articles.



Exercice 5 : Premier précalcul

□ **Étape 1** : Extraction des thématiques du corpus.

