

Présentation générale de la base ISTE^X

Journée ISTE^X - 19 Janvier 2021

Au programme

INTRODUCTION

- ▣ **ISTEX** : présentation & actualité

FOCUS SUR QUELQUES SERVICES

- ▣ **DATA.ISTEX** : une autre vue sur les données ISTEX
- ▣ **ISTEX-DL** : télécharger un corpus ISTEX en quelques clics

1.

Introduction

ISTEX

L'excellence documentaire pour tous



Construire le socle de la bibliothèque scientifique numérique nationale.

- 2011 - 2018, un projet créé dans le cadre des PIA
- Aujourd'hui, un service pour l'ESR**



Présentation

- Acquisition massive et centralisée d'archives scientifiques :
collections rétrospectives de la littérature scientifique dans toutes les disciplines
- Mise à disposition et signalement

Mode d'accès

- Réservé à l'enseignement supérieur et la recherche.
- Accessible par adhésion

356 établissements

ISTEX

| Authentification

Vous êtes sur le point de lancer l'adhésion à ISTEX, si vous voulez vous informer sur ce qu'offre l'adhésion, cliquez [ici](#).

L'identifiant et le mot de passe à utiliser sont ceux du site [licencesnationales.fr](#)

Identifiant	<input type="text" value="identifiant"/>
Mot de passe	<input type="password" value="mot de passe"/>

[Se connecter](#)

[Vous avez oublié votre mot de passe ?](#)

Votre établissement n'a pas encore de compte ? Vous serez dirigé sur le site [licencesnationales.fr](#) de l'ABES pour en créer un.

[+ Créer un compte](#)







© 2018 ABES [Nous Contacter](#) [CGU](#) - [DONNÉES PERSONNELLES](#) - [MENTIONS LÉGALES](#)

Une plateforme



1. Usage documentaire



- api.istex.fr 
- Bouton 
- Google Scholar 
- Outils Biblio. 



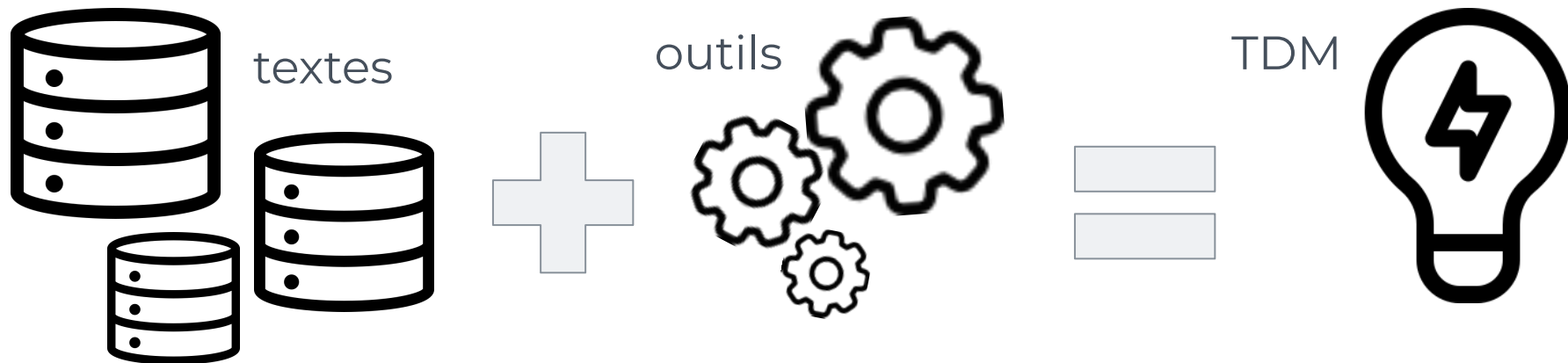
Enrichissement

2. Usage TDM

-  data.istex.fr
-  dl.istex.fr
-  corpus.istex.fr

Gargantext, Cortext, Iramuteq ...

Adaptée pour la fouille et l'analyse



... des nouvelles perspectives
... des nouvelles données de
recherche

De nombreux avantages

- **Un seul lieu** pour des nombreuses sources
- Un accès **sans restriction** à des données standardisées, **interopérables** et **enrichies**
- Un **cadre juridique** sécurisé : une licence appropriée
- **Des connexions** vers des outils/plateformes créés par le monde académique

Des données de qualité

Protocole de réception des données reçus :

- ★ Analyse d'échantillon & réception de corpus
- ★ Analyse quantitative et qualitative des métadonnées
- ★ Analyse de la cartographie des fichiers reçus
- ★ Validation avant mise en ligne

✔ Un signalement des métadonnées de qualité

👍 **Pour accéder aux documents**

Des données interoperables

- ★ Homogénéisation (langues, type de documents, etc.)
 - ★ Correction des erreurs
 - ★ Transformation dans des formats standards :
TEI, MODS
-
- ✔ Gommer les différences entre fournisseurs
 - 👍 **Minimiser les “pré-traitements” pour le TDM**

Des données enrichies

et en exclusivité

données
d'origine



données originales

+

nouvelles données

+

nombreux critères de recherche

valeur ajoutée

(ré) Océrisation

676

Intraventricular kainic acid preferentially destroys hippocampal pyramidal cells

THE hippocampus is particularly vulnerable to a variety of conditions, such as anoxia, status epilepticus and senile dementia, in which central neurones are lost^{1,2}. Most commonly, the lesion involves only the Sommer sector (h₁) and the endfolium (h₃-h₅), sparing area h₂, the fascia dentata and most regions outside the hippocampal formation. The consequences for hippocampal connections are unknown. Studies on the rat hippocampus suggest that connections made by the affected neurones could be replaced by axons of other neurones which project to the same areas^{3,4}. These anomalous synapses might either compensate in part for the loss of cells or contribute to whatever functional deficits may derive from the lesion. Since a good deal is known about afferent and efferent hippocampal connections in the rat, this animal might serve as a model for studies of hippocampal damage. However, the selective pathology seen clinically cannot be reproduced by conventional lesioning techniques. Ideally, one would like to use a toxin relatively specific for the neurones in question. Kainic acid, a potent excitatory analogue of glutamic acid^{5,7}, has been used to destroy neurones in the arcuate nucleus⁸ and striatum,⁹⁻¹¹ while sparing fibres which pass to or through these regions. Previous workers have also briefly noted lesions in the hippocampus,^{8,11} but these were not described. Accordingly, we injected kainic acid intraventricularly into the rat brain and studied its effect on hippocampal neurones. We now report the unusual sensitivity of CA3-CA4, and to a lesser extent CA1, pyramidal cells to this agent. Our results suggest that kainic acid lesions can provide a model of hippocampal damage in man.



676

_ I

Intraventricular kainic acid preferentially destroys hippocampal pyramidal cells



THE hippocampus is particularly vulnerable to a variety of conditions, such as anoxia, status epilepticus and senile dementia, in which central neurones are lost^{1,2}. Most commonly, the lesion involves only the Sommer sector (h) and the endfolium (h₃-h₅), sparing area h₂, the fascia dentata and most regions outside the hippocampal formation. The consequences for hippocampal connections are unknown. Studies on the rat hippocampus suggest that connections made by the affected neurones could be replaced by axons of other neurones which project to the same areas^{3,4}. These anomalous synapses might either compensate in part for the loss of cells or contribute to whatever functional deficits may derive from the lesion. Since a good deal is known about afferent and efferent hippocampal connections in the rat, this animal might serve as a model for studies of hippocampal damage. However, the selective pathology seen clinically cannot be reproduced by conventional lesioning techniques. Ideally, one would like to use a toxin relatively specific for the neurones in question. Kainic acid, a potent excitatory analogue of glutamic acid⁵, has been used to destroy neurones in the arcuate nucleus⁸ and striatum⁹⁻¹¹ while sparing fibres which pass to or through these regions. Previous workers have also briefly noted lesions in the hippocampus^{8,11} but these were not described. Accordingly, we injected kainic acid intraventricularly into the rat brain and studied its effect on hippocampal neurones. We now report the unusual sensitivity of CA3-CA4, and to a lesser extent CA1, pyramidal cells to this agent. Our results suggest that kainic acid lesions can provide a model of hippocampal damage in man.

Caractérisation des textes

- Score de qualité
- Qualité des PDF
- Nombre de mots
- Présence et type d'enrichissements

Eur J Med Sci (2010) 179:259–263
DOI 10.1007/s11845-009-0432-3

ORIGINAL ARTICLE

The cervical spine of professional front-row rugby players: correlation between degenerative changes and symptoms

B. A. Hogan · N. A. Hogan · P. M. Vos ·
S. J. Eustace · P. J. Kenny

Received: 6 October 2008 / Accepted: 14 September 2009 / Published online: 8 October 2009
© Royal Academy of Medicine in Ireland 2009

Abstract

Background Injuries to the cervical spine (C-spine) are among the most serious in rugby and are well documented. Front-row players are particularly at risk due to repetitive high-intensity collisions in the scrum.

Aim This study evaluates degenerative changes of the C-spine and associated symptomatology in front-row rugby players.

Materials and methods C-spine radiographs from 14 professional rugby players and controls were compared. Players averaged 23 years of playing competitive rugby. Two consultant radiologists performed a blind review of radiographs evaluating degeneration of disc spaces and apophyseal joints. Clinical status was assessed using a

modified AAOS/NASS/COSS cervical spine outcomes questionnaire.

Results Front-row rugby players exhibited significant radiographic evidence of C-spine degenerative changes compared to the non-rugby playing controls ($P < 0.005$). Despite these findings the rugby players did not exhibit increased symptoms.

Conclusion This highlights the radiologic degenerative changes of the C-spine of front-row rugby players. However, these changes do not manifest themselves clinically or affect activities of daily living.

Keywords Rugby · Cervical spine · Degenerative change · Front-row

B. A. Hogan (✉)

Department of Diagnostic Imaging, Sports Surgery Clinic,
Sandy Demesne, Dublin 9, Ireland
e-mail: bhogan@eircom.net

N. A. Hogan

Department of Orthopaedic Surgery,
Sports Surgery Clinic, Dublin, Ireland

P. M. Vos

Department of Radiology,
St. Paul's Hospital, Vancouver, BC, Canada

N. A. Hogan · P. J. Kenny

Department of Orthopaedic Surgery,
Cappagh National Orthopaedic Hospital,
Dublin, Ireland

S. J. Eustace

Department of Radiology,
Cappagh National Orthopaedic Hospital,
Dublin, Ireland

Introduction

Injuries to the cervical spine (C-spine) are among the most serious injuries occurring in rugby [1]. The earliest published reference to the relationship between rugby and spinal injuries dates back to a report in *The Times* of London from November 1871, in which it was stated that "the most serious injuries sustained by players in the scrum is a phase method of re-starting between the two 'Hi' and is used as a the opposing pack. This the phase of play most injuries [1, 4, 8].

While some studies have reported the incidence of injuries to be higher among adults



Structuration des PDF

Identifier le titre, le résumé, les paragraphes des articles

Automatic Extraction and Resolution of Bibliographical References in Patent Documents

Patrice Lopez
patrice.lopez@hotmail.com

Abstract. This paper describes experiments with Conditional Random Fields (CRF) for extracting bibliographical references in patent documents. CRF are used for performing extraction and parsing tasks which are expressed as sequence tagging problems. The automatic recognition covers references to other patent documents and to scholarship publications which are both characterized by a strong variability of contexts and patterns. Our work is not limited to the extraction of reference blocks but also includes fine-grained parsing and the resolution of the bibliographical references based on data normalization and the access to different online bibliographical services. For these different tasks, CRF models surpass significantly existing rule-based algorithms and other machine learning techniques, resulting more particularly in a very high performance for patent reference extractions with a reduction of approx. 75% of the error rate compared to previous works.

Introduction

Bibliographical citations play a major role in patent information. Citations report the closest prior art which will be the basis for evaluating the contribution a patent application and for identifying grantable subject matter. In patent cases, the result of the search phase is the *search report*, a collection of references to patents and to other public documents such as scientific articles, technical manuals or research disclosures, so-called Non-Patent Literature (NPL). In addition to the search report, the text body of the patent document contains usually many bibliographical references introduced in the original application documents or introduced at a further filing stage or at granting stage. A patent

```
<?xml version="1.0" encoding="UTF-8" ?>
<document>
  <title level="1" type="Section-Header">Automatic Extraction and Resolution of Bibliographical References in Patent Documents</title>
  <author level="1" type="Text">Patrice Lopez</author>
  <abstract level="1" type="Text">This paper describes experiments with Conditional Random Fields (CRF) for extracting bibliographical references in patent documents. CRF are used for performing extraction and parsing tasks which are expressed as sequence tagging problems. The automatic recognition covers references to other patent documents and to scholarship publications which are both characterized by a strong variability of contexts and patterns. Our work is not limited to the extraction of reference blocks but also includes fine-grained parsing and the resolution of the bibliographical references based on data normalization and the access to different online bibliographical services. For these different tasks, CRF models surpass significantly existing rule-based algorithms and other machine learning techniques, resulting more particularly in a very high performance for patent reference extractions with a reduction of approx. 75% of the error rate compared to previous works.</abstract>
  <introduction level="1" type="Text">Bibliographical citations play a major role in patent information. Citations report the closest prior art which will be the basis for evaluating the contribution a patent application and for identifying grantable subject matter. In patent cases, the result of the search phase is the search report, a collection of references to patents and to other public documents such as scientific articles, technical manuals or research disclosures, so-called Non-Patent Literature (NPL). In addition to the search report, the text body of the patent document contains usually many bibliographical references introduced in the original application documents or introduced at a further filing stage or at granting stage. A patent</introduction>
  </document>
</pre>
```

Automatic Extraction and Resolution of Bibliographical References in Patent Documents

Patrice Lopez
patrice.lopez@hotmail.com

Abstract. This paper describes experiments with Conditional Random Fields (CRF) for extracting bibliographical references in patent documents. CRF are used for performing extraction and parsing tasks which are expressed as sequence tagging problems. The automatic recognition covers references to other patent documents and to scholarship publications which are both characterized by a strong variability of contexts and patterns. Our work is not limited to the extraction of reference blocks but also includes fine-grained parsing and the resolution of the bibliographical references based on data normalization and the access to different online bibliographical services. For these different tasks, CRF models surpass significantly existing rule-based algorithms and other machine learning techniques, resulting more particularly in a very high performance for patent reference extractions with a reduction of approx. 75% of the error rate compared to previous works.

Introduction

Bibliographical citations play a major role in patent information. Citations report the closest prior art which will be the basis for evaluating the contribution a patent application and for identifying grantable subject matter. In patent cases, the result of the search phase is the *search report*, a collection of references to patents and to other public documents such as scientific articles, technical manuals or research disclosures, so-called Non-Patent Literature (NPL). In addition to the search report, the text body of the patent document contains usually many bibliographical references introduced in the original application

Extraction des références bib.

Détecter et structurer
les références
bibliographiques des
articles en XML TEI

References

1. Lopez, P., Romary, L.: Multiple retrieval models and regression models for prior art search. In: CLEF 2009 Workshop, Technical Notes, Corfu, Greece (2009)
2. Nakov, P., Schwartz, A., Hearst, M.: Citances: Citation sentences for semantic



```
▼<biblStruct xml:id="b0" resp="#ISTEX-API" change="#refBibs-istex">
  ▼<analytic>
    ▼<title level="a" type="main">
      Multiple retrieval models and regression models for prior art search
    </title>
    ▼<author>
      ▼<persName>
        <forename type="first">P</forename>
        <surname>Lopez</surname>
      </persName>
    </author>
    ▼<author>
      ▼<persName>
        <forename type="first">L</forename>
        <surname>Romary</surname>
      </persName>
    </author>
  </analytic>
  ▼<monogr>
    <title level="m">CLEF 2009 Workshop</title>
    ▼<meeting>
      ▼<address>
        <addrLine>Corfu, Greece</addrLine>
      </address>
    </meeting>
    ▼<imprint>
      <date type="published" when="2009"/>
    </imprint>
  </monogr>
</biblStruct>
```

Catégorisation des documents

- Par appariement :
WoS, 
Scopus, **Scopus***
Science Metrix 
Science-Metrix
- Par apprentissage automatique :
Classification Pascal/Francis 

JOURNAL OF Plant Physiology
© 1997 by Gustav Fischer Verlag, Jena

Concentration of Zinc and Activity of Copper/Zinc-Superoxide Dismutase in Leaves of Rye and Wheat Cultivars Differing in Sensitivity to Zinc Deficiency

I. CAKMAK¹*, L. ÖZTÜRK¹, S. EKER¹, B. TORUN¹, H. I. KALFA¹, and A. YILMAZ²
¹ Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Cukurova University Adana, Turkey
² International Winter Cereals Research Center, POB 325 Konya, Turkey
Received July 16, 1996 - Accepted October 30, 1996

Summary
Two bread wheat (*Triticum aestivum* L. cv. Besotaj-1 and BDME-10), two durum wheat (*Triticum durum* L. cv. Karaduru-1149 and Karilata-91) and one rye (*Cereale cernale* L. cv. Adana) cultivars differing in sensitivity to zinc (Zn) deficiency were grown in a Zn deficient soil to compare severity of Zn deficiency of total superoxide dismutase (SOD), copper/zinc (Cu/Zn)-SOD, and manganese (Mn)-SOD (Mn-SOD) in leaves. Visual Zn deficiency symptoms such as chlorosis appeared rapidly and were severe in bread wheat cultivars, while Besotaj-1 was much less affected and either absent or only slightly developed. The Mn-SOD activity was similar to the effect on leaf symptoms. Cu/Zn-SOD activity was about 16% in Adana (rye) and Besotaj-1. Despite of such marked differences in sensitivity to Zn deficiency between the cultivars, Cu/Zn-SOD activity was not different between the cultivars under Zn deficiency. Under Zn deficiency, rye showing a high Cu/Zn-SOD. Among the wheat cultivars, Besotaj-1 was the most sensitive to Zn deficiency. The results suggested that Zn efficiency in cereals is determined by the expression of Zn concentration in leaves and that activity of Cu/Zn-SOD is a Zn concentration index. An efficient utilization of Zn in leaves is a determining expression of Zn efficiency in cereals.

Key words: *Cereale cernale*, *Triticum aestivum*, *Triticum durum*, superoxide dismutase, zinc concentration, zinc deficiency, zinc efficiency

Catégorisation par appariement
WoS : Plant Sciences

Catégorisation par apprentissage
Agronomie, Sciences du sol et productions végétales

Indexation automatique

Extraire du texte les termes les plus représentatifs du contenu quel que soit le domaine scientifique

A Retrospective Mortality Study of Workers Exposed to Arsenic in a Gold Mine and Refinery in France

L. Simonato, MD, J.J. Moulin, MD, B. Javelaud, MD, PhD, G. Ferro, BSc, P. Wild, BSc, R. Winkelmann, MA, and R. Saraccl, MD

A historical mortality study of a cohort of employees of a gold mining and refining company was carried out in Salsigne, France. A major goal of the study was to investigate the relationship between lung cancer mortality and exposure to arsenic, radon, silica, and other contaminants of the working environment. A twofold excess of lung cancer was found both among miners and smelters, mainly concentrated among workers who had experienced exposure to past levels of arsenic, radon, and silica. The consistency of the results in the mine and the refinery are suggestive of a carcinogenic risk from both soluble and insoluble arsenic, although the potential role of other factors cannot be dismissed. © 1994 Wiley-Liss, Inc.

Key words: radon, silica, gold mining and refining, retrospective cohort, lung cancer

INTRODUCTION

An apparent high incidence of neoplasms of the respiratory system among employees in gold extraction and refining in Salsigne (Aude) was first reported in 1977 [doctoral thesis by Perisse, 1976-77] from the Department of Pneumology of the General Hospital in Carcassonne. Forty cases of lung cancer were included in the first investigation, whose results, even in the absence of a formal comparison group, appeared to indicate a large excess when considering the time period and the size of the population studied. A similar case series was subsequently reported in 1985 in another doctoral thesis written by Jammes [1985].

```
<listAnnotation type="rd-tweet">
  <annotationBlock corresp="text" xmlns="https://www.tei-c.org/ns/1.0">
    <keywords change="#istes-rd" resp="#istes-rd">
      <term>
        <term lung cancer /term>
        <fs type="statistics">
          <ef name="frequency">
            <numeric value="17"/>
          </ef>
          <ef name="specificity">
            <numeric value="1"/>
          </ef>
        </fs>
      </term>
      <term>
        <term radon /term>
        <fs type="statistics">
          <ef name="frequency">
            <numeric value="14"/>
          </ef>
          <ef name="specificity">
            <numeric value="0.823529411764706"/>
          </ef>
        </fs>
      </term>
    </keywords>
  </annotationBlock>
</listAnnotation>
```

- Lung Cancer
- Cohort
- Arsenic
- Miner
- Refinery
- Salsigne
- Diesel exhaust
- First exposure

Détection des entités nommées

9 types d'entités :

- Personnes,
- Lieux,
- Organisations,
- Projets financés,
- Organisme financeur,
- Hébergeur de ressources,
- URL,
- Dates,
- Citations

INTRODUCTION

An apparent high incidence of neoplasms of the respiratory system among employees in gold extraction and refining. **Salsigne** (Aude) was first reported in 1977 [doctoral thesis by Perisse, 1976-77] from the **Department of Pneumology of the General Hospital in Carcassonne**. Forty cases of lung cancer were included in the first investigation, whose results, even in the absence of a formal comparison group, appeared to indicate a large excess when considering the time period and the size of the population studied. A similar case series was subsequently reported in **1985** in another doctoral thesis written by Jammes [1985].

Désambiguïisation des lieux

Alignement des entités nommées de type PlaceName avec la ressource GeoNames

The current issue and full text archive of this journal is available at www.emeraldinsight.com/0025-1747.htm

MD
47,1

24

Management implications of foreign players in the English Premiership League football

Nnamdi Madichie
University of East London Business School, London, UK

Position	Team	P	W	D	L	F	A	GD	Pts
1	Manchester United	38	27	6	5	80	22	58	87
2	Chelsea	38	25	10	3	65	26	38	85
3	Arsenal	38	24	11	3	74	31	43	83
4	Liverpool	38	21	13	4	67	28	39	76
5	Everton	38	19	8	11	55	33	22	65
6	Aston Villa	38	16	12	10	51	51	0	60
7	Manchester City	38	15	14	9	50	48	2	58
8	Portsmouth	38	16	9	13	48	40	8	57
9	Manchester City	38	15	10	13	45	53	-8	55
10	West Ham United	38	13	10	15	42	50	-8	49

Source: www.skysports.com/football/league/0,19540,11660,00.html

Table III
Top ten Barclays
Premiership teams (end
of season - 11 May 2008)



Portsmouth

FEDERAL EMPLOYEES AND VOTING IN FEDERAL ELECTIONS

DALMAS H. NELSON, *San Fernando Valley State College*
and
PAUL J. HOFFMANN, *St. Paul, Minnesota*

AMONG the many uncertainties of American political behavior are the voting practices of those who work for the federal government. Though they constitute a substantial number of voters in the nation and their support is cultivated somewhat by both major political parties, little solid information has been available as to their ways at the polls. This paper examines their role in federal elections. Speculation about federal employee voting practices has tended to take divergent directions. One view is that represented by *Business Week*, which says that federal employees have a stake in governmental activity and large expenditures which "generally make them Democrats in national politics, even during Republican administrations."³

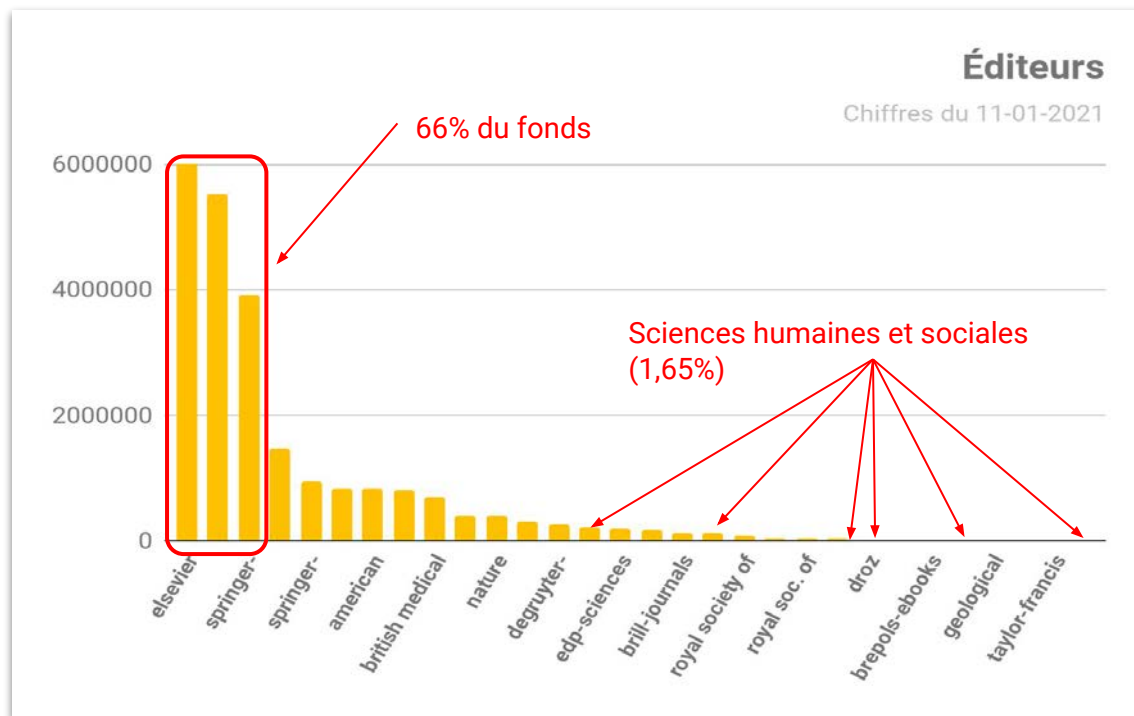
The third group is made up of the three Virginia cities of Alexandria, Falls Church, and Portsmouth. They are, on the map, within the perimeters of Fairfax County in the case of Alexandria and Falls Church, and of Norfolk County in the case of Portsmouth.⁴ But technically they are outside of the counties by virtue of city-county separation. All three are Section 16 cities, Alexandria and Falls Church due to their proximity to the capital and Portsmouth as the result of determination by the Commission that federal personnel (and their wives) constitute a majority of the voters. The three separated cities are grouped together because the electoral process in these separated cities was felt to be significantly different from that of either counties or municipalities not separated from their counties.

Le fonds ISTEX

Les principaux éditeurs scientifiques

Elsevier, Wiley et Springer journals totalisent 66%

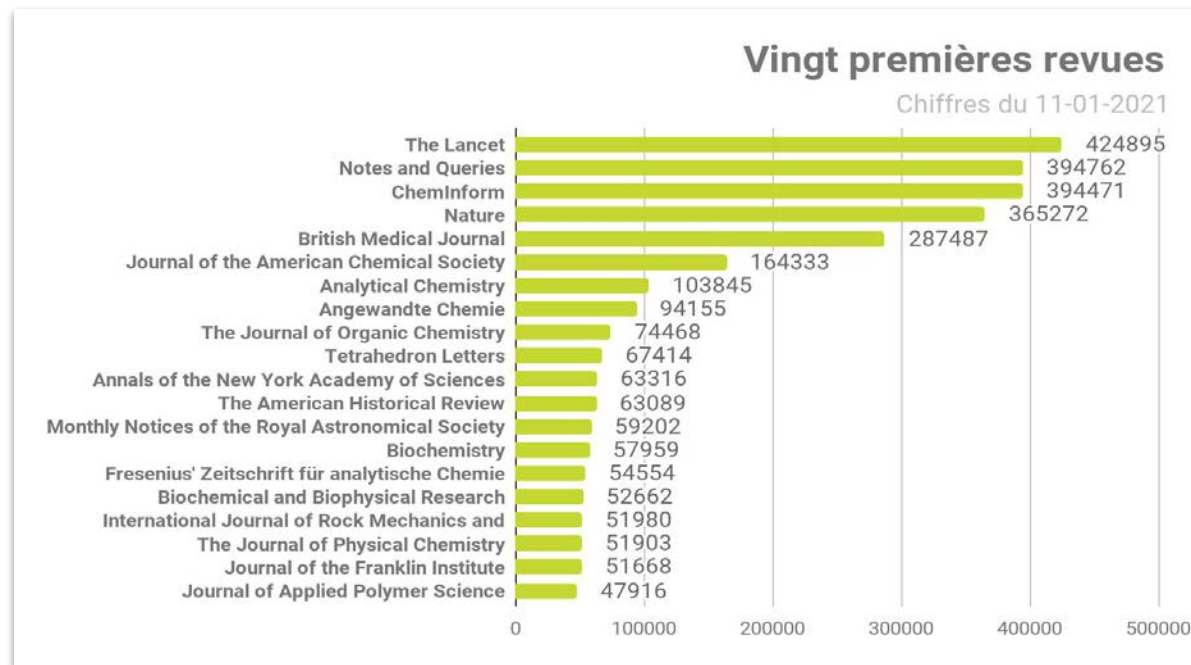
6 éditeurs spécialisés en SHS représentent 1.65% (mais disciplines également présentes chez d'autres éditeurs)



Les plus grandes revues scientifiques

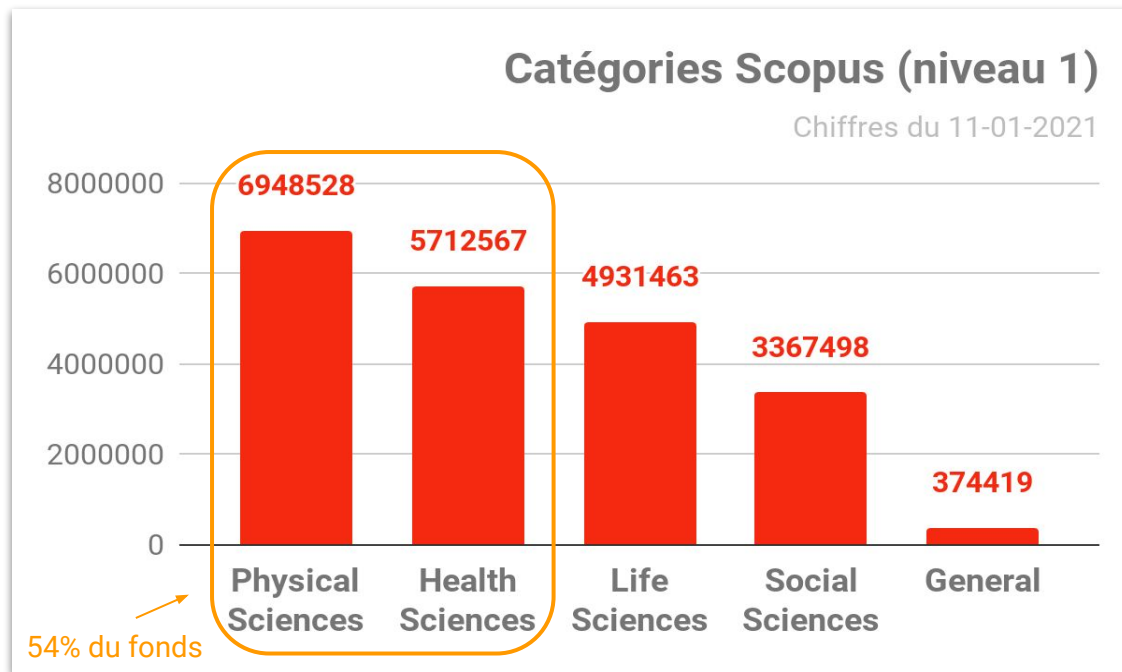
Dans le fonds de plus de **9 000** revues présentes dans ISTE^X :

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



Tous les domaines scientifiques

54% font partie des sciences physiques ou de la santé



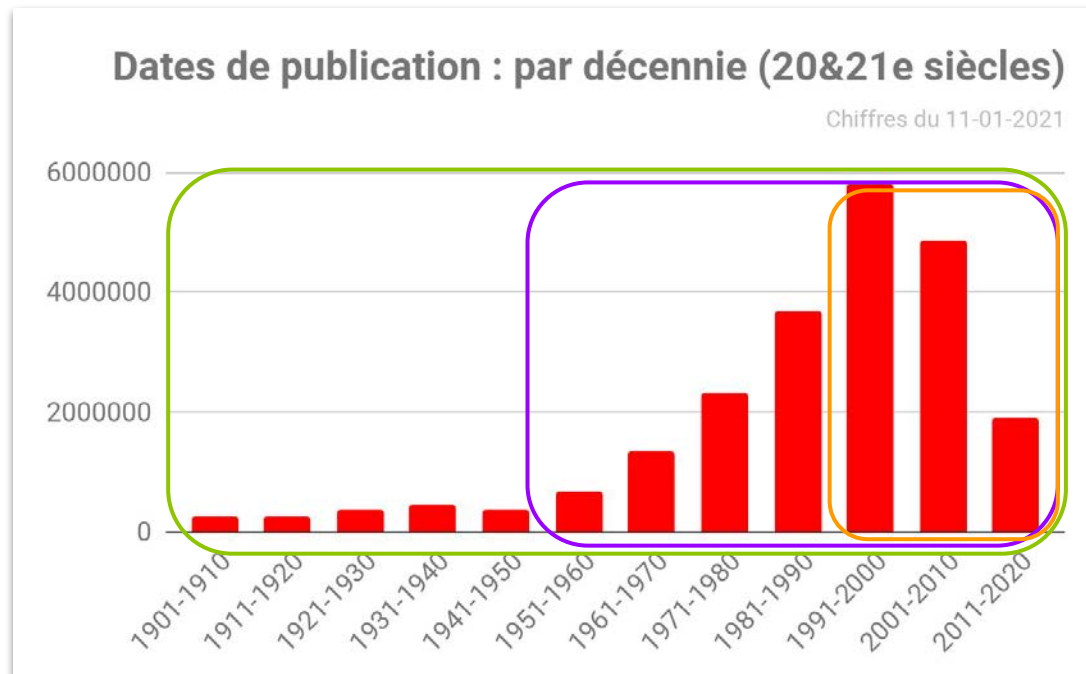
700 ans de publication

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

88% des documents publiés
depuis 1950

54% des documents publiés
sur les 30 dernières années

5% des documents publiés
avant 1900

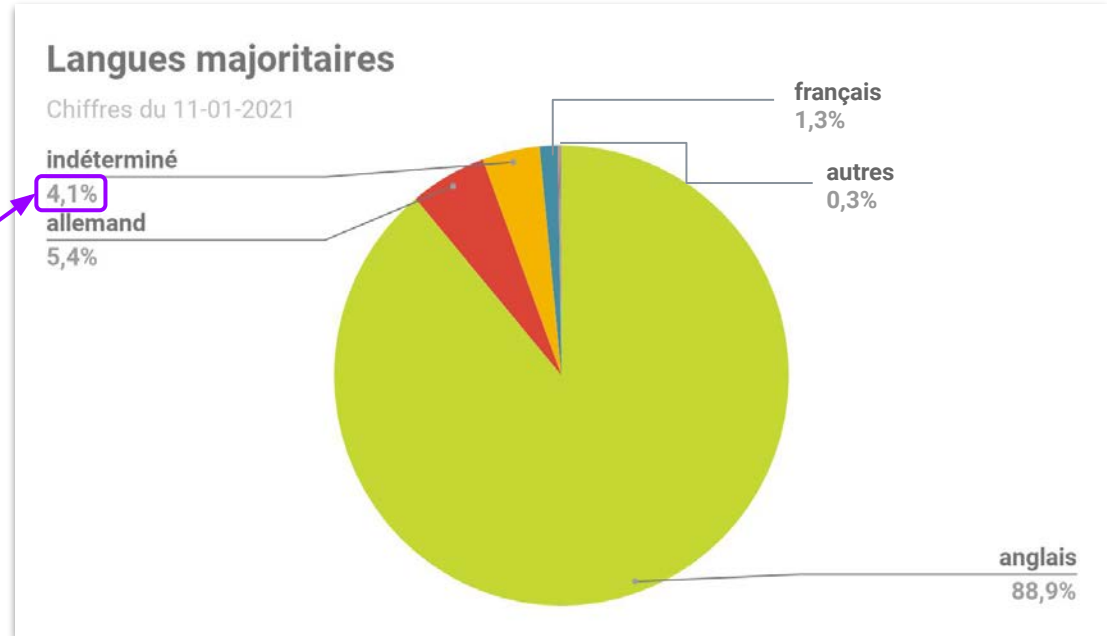


Polyglotte : 52 langues !

Anglais majoritaire

0,3% = 48 autres langues

Information non renseignée par les éditeurs pour près de 1 million de documents !



Résumé en chiffres

- 30 corpus éditeur
- **9 307** revues
- 38 858 monographies
 - 348 056 avec EEBO & ECCO
- **23 351 350** documents
- **98%** de documents enrichis
- **~50 000** accès par semaine
 - soit ~4000 PDF téléchargés
 - dont 40% d'accès documentaire

**En constante
évolution**

Analyse en cours

- Cambridge University Press
110 ebooks en science politique
- De Gruyter
1.291 ebooks et numéros de revues en sciences de l'Antiquité
- Brill
240 ebooks en religion grecque et romaine
- Karger
133 ebooks en médecine humaine

- Proquest
80.700 archives du Parlement britannique du IXe siècle
- Proquest
ebooks du Twentieth Century Religious Thought – Christianity
- Proquest
2750 ebooks de Early European Books – Bibliothèque nationale de Florence
- CAIRN
13 revues françaises

Mise à jour à venir

Augmentation de la couverture pour


- EDP Sciences
 - jusqu'à 2019 → + 12 079 documents, 28 titres
 - ensuite : 2020, 2021
- Elsevier
 - 2002 à 2008 → +2,2 millions d'articles
 - ensuite : 2009, 2010, 2011, 2012.

...

Pour aller plus loin...

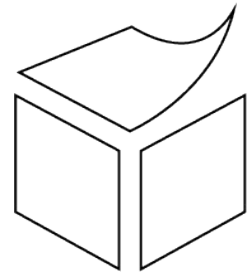


Plusieurs sites accessibles depuis www.istex.fr

 en 2021 réorganisation du site pour améliorer son expérience utilisateur

2.

data.istex



Une autre vision sur les données ISTEK

- ▣ Pour la fouille de données
 - ▣ Des exemples de corpus spécialisés répondant à différents cas d'usage
 - ▣ Des enrichissements provenant de traitements TDM
- ▣ Pour le web de données
 - ▣ Des données produites et/ou transformées à l'Inist selon les normes du web sémantique
 - ▣ Des jeux de données interrogeables par le triple store

Comment y accéder ?

<https://data.istex.fr>

23 millions de documents
littérature scientifique
9 307 revues et 348 000 articles

Testez ISTEEX : indiquez un titre, des mots-clés ou un DOI



Bouton



Scholar



Zotero



Télécharger



API



Harvester



SPARQL



data.istex.fr



Rechercher

Les différentes vues d'ISTEX

les états de collection des revues présentes dans ISTE

Revue De Sommaire

L'édition ISTE



+ Présentation S. Gregorio à 12h20

métadonnées et enrichissements des documents ISTE

Référentiels Documentaires

En savoir plus sur les métadonnées des ressources ISTE

+ EN SAVOIR PLUS

Corpus Scientifiques

Morceaux choisis d'ISTE

documents rassemblés autour d'une thématique scientifique ou d'un sujet d'actualité

+ EN SAVOIR PLUS

Usage Istex

Au service des utilisateurs

Qui se sert d'ISTE et comment

+ EN SAVOIR PLUS

Focus sur ...



Les corpus scientifiques

Corpus Actualité

Explorer le passé pour éclairer le présent



EN SAVOIR PLUS

NEW

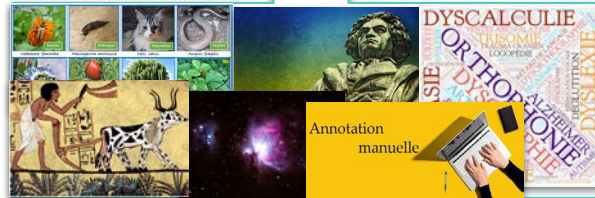


Corpus Spécialisés

Des collections de corpus destinés à la fouille de texte



EN SAVOIR PLUS



Publications Remarquables

Des textes et des auteurs illustres



EN SAVOIR PLUS



Des exemples de corpus spécialisés

▣ Animalia 100

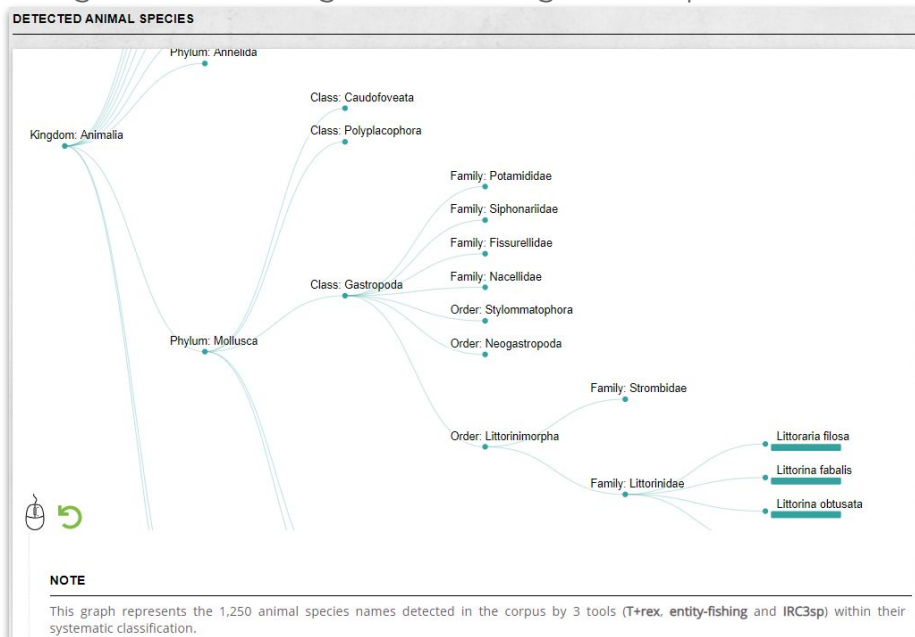
- ▣ Collection “Systématique”
- ▣ Corpus de documents en zoologie
- ▣ Corpus enrichi avec l’annotation d’entités nommées scientifiques
 - Détection automatique des noms d’espèces animales par 3 outils
 - Ajout de leur classification systématique
 - Mise à disposition des annotations automatiques
- ▣ Article présenté à **LREC 2020** (12th International Conference on Language Resources and Evaluation)
 - “An experiment in annotating animal species names from ISTEX resources”
(S. Barreaux, D. Besagni)




Des exemples de corpus spécialisés

ANIMALIA 100

Navigation dans le règne animal au gré des espèces trouvées



Pour chaque document du corpus, liste des noms d'espèces trouvés dans le texte

DOCUMENT TITLE	
Dual-energy X-ray Absorptiometry of Birds: an Examination of Excised Skeletal Specimens	
CORPUS NAME	LINK TO THE DOCUMENT
Animalia 100	
DETECTED ANIMAL SPECIES NAMES	
<ul style="list-style-type: none">• Bonasa umbellus• Colinus virginianus• Gallus domesticus• Meleagris gallopavo	

Des exemples de corpus spécialisés

ANIMALIA 100

A venir :

- Des pages dédiées aux espèces animales pour afficher des informations complémentaires
 - Classification systématique
 - Liens vers bases de référence (Catalogue of Life, Wikidata)
- Une navigation plus poussée

À suivre !

Des exemples de corpus spécialisés

▣ Ludwig van Beethoven

- ▣ Collection “Beethoven”
- ▣ Corpus de 1 800 documents en musicologie
- ▣ Corpus à objectif pédagogique
 - Illustre une méthode de constitution de corpus pour une application en fouille de textes
 - En exploitant les enrichissements produits par l'INIST
- ▣ Démo présentée à **TALN 2020** (Traitement Automatique des Langues Naturelles)
 - “Vers un corpus optimal pour la fouille de textes : stratégie de constitution de corpus spécialisés à partir d'ISTEX” (C. de Salabert, S. Barreaux)



Démo C. de
Salabert à 16h00

Des exemples de corpus spécialisés

BEETHOVEN

Les entités nommées du corpus

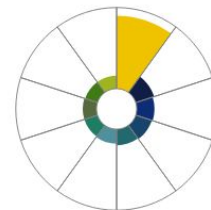


Les entités nommées d'un document avec recherche des documents similaires

ENTITÉS NOMMÉES "NOM DE PERSONNE" (UNITEX)

Richard Strauss, Wilhelm Broel, Max Unger, Hugo von Hofmannsthal, Stephen Ley, Willy Hess, Beethoven, Hans von BiiLOW, Willi Schuh, Mies, Friedrich Munter, M. M. S. Beethoven, Philipp Losch

PUBLICATIONS SIMILAIRES (ENTITÉS NOMMÉES)



NOTE

Représentation des dix publications ayant le plus d'entités nommées de type "nom de personne" en commun avec cette ressource

Des exemples de corpus spécialisés

▣ EN-ISTEX

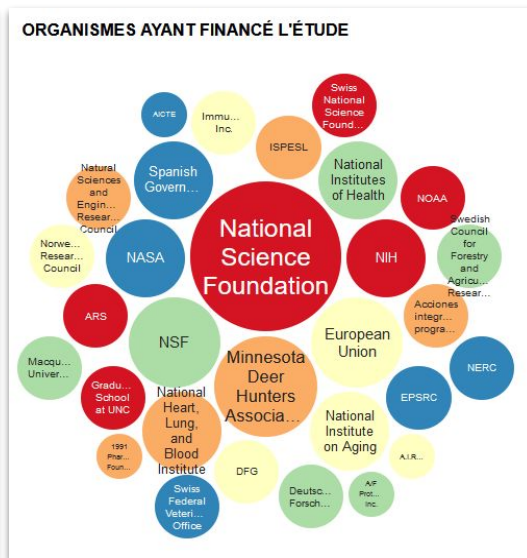
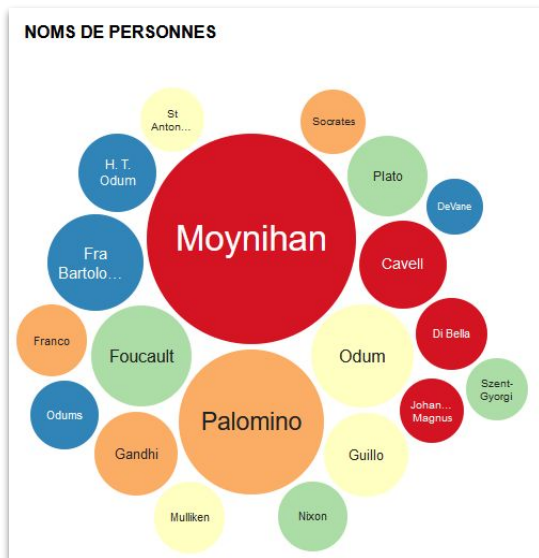
- ▣ Collection “Corpus annotés manuellement”
- ▣ Corpus multi-domaines
- ▣ Corpus enrichi avec l’annotation d’entités nommées
 - Entités nommées ISTEEX
 - Fiabilité mesurée par un **accord inter-annotateur**
 - Mise à disposition du **guide d’annotation** & des **annotations manuelles** afin de tester des outils de reconnaissance d’entités nommées
- ▣ Article soumis à TALN 2021 (Traitement Automatique des Langues Naturelles)
 - *“Corpus EN-ISTEX : un corpus d’articles scientifiques annoté manuellement en entités nommées”*



Des exemples de corpus spécialisés

EN-ISTEX

Les entités nommées du corpus



Les entités nommées d'un document

TITRE DE L'ARTICLE

The Italian guidelines for early intervention in schizophrenia: development and conclusions

LIEN VERS LE PDF



PERSNAME

- Corrado Barbui
- Giovanni Neri
- Angelo Picardi
- Andrea Alpi
- Silvia Grignani
- Rosaria Rosanna Cammarano
- Mario Maj
- Vincenzo Pastore
- Michele Procacci
- Michele Tansella
- Paolo Brambilla
- Mirella Ruggeri
- Domenico Semisa
- Antonio Lora

PLACENAME

- Melbourne
- Australia
- Norway
- Rogaland County
- Norway
- London
- Ontario
- Canada

DATE

- from September 2004 to May 2007
- from January 2000 to June 2006

Les différentes vues d'ISTEX



Focus sur ...



L'usage d'ISTEX



▣ Publications sur ISTEX

- ▣ Le seul jeu de données qui n'expose pas des publications issues d'ISTEX, mais des publications **sur** ISTEX :
 - Par équipe INIST : sur le développement des services et enrichissements de données
 - Par équipes scientifiques : sur usage ISTEX et ses corpus en texte intégral comme données de recherche pour fouille de texte

Focus sur “usage ISTEEX”

PUBLICATIONS SUR ISTEEX

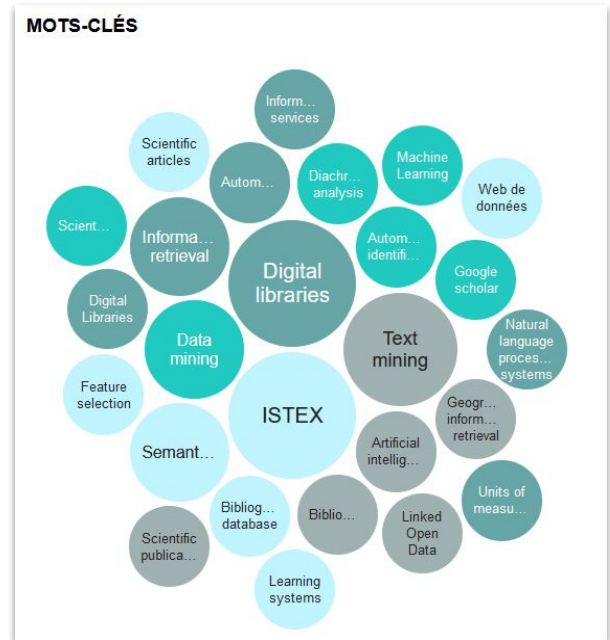
La liste des ressources en mode recherche et filtrage par facettes

48 ressources trouvées sur un total de 48

- Auteur(s) (91)
- Mots-clés (184)
- Langue (4)
- Type document (9)
- Année (9)
- Financements Détails (19)

TITLE
Sept étapes pour publier des données ouvertes et liées
Vers un corpus optimal pour la fouille de textes : stratégie de constitution de corpus spécialisés à partir d'ISTEEX
A Machine Learning System for Assisting Neophyte Researchers in Digital Libraries
L'API Istex : le sésame pour accéder aux ressources acquises
Les services Istex. Istex au-delà de l'acquisition
Is the Abstract a Mere Teaser? Evaluating Generosity of Article Abstracts in the Environmental Sciences
The Bibliothèque scientifique numérique (BSN, digital science library): Providing a political framework for STI co-ordination
Evaluating the Usefulness of Citation Graph and Document Metadata in Scientific Document Recommendation for Neophytes
ISTEX : DES ENRICHISSEMENTS AU WEB DE DONNÉES

Les mots-clés d'auteur du corpus



L'envers du décor



En évolution constante !

- ▣ Application web open source dédiée aux données structurées
 - ▣ Exposition sur le web
 - ▣ Visualisation grâce à des graphiques
 - ▣ Export selon les normes du web sémantique
- ▣ Une navigation facilitée :
 - ▣ Fil d'Ariane
 - ▣ Interface responsive compatible avec mobiles
- ▣ Une exploitation simplifiée
 - ▣ Bouton "Télécharger" directement connecté à ISTEEX-DL

L'envers du décor

data.istex.fr

data.istex.fr expose le Triple Store des données ISTEI via son SPARQL Endpoint.

← Data ISTEIX / Corpus scientifiques / Corpus spécialisés / Collection Beethoven

Fil d'Ariane

Corpus Ludwig van Beethoven

Documents de musicologie issus d'ISTEX relatifs au compositeur Beethoven

VERSION	NOMBRE DE DOCUMENTS
VI	1864

DESCRIPTION DU CORPUS

Télécharger

Accueil Graphiques Recherche



https://dl.istex.fr/?withID=true&q_id=4c8246b6926fo4dc11a0ed76209f375b&extract=&size=1864&rankBy=qualityOverRele...

ISTEX

Téléchargez un corpus ISTEIX

Vous êtes membre de l'Enseignement supérieur et de la Recherche et vous souhaitez extraire un corpus de documents ISTEIX ?
3 étapes suffisent pour récupérer une archive compressée de votre corpus sur votre disque dur.

1. Requête

Explicititez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous :

Équation booléenne Identifiants ARK Import de fichier Exemples

ark:/67375/M70-S1JGFLBL-5

L'équation saisie correspond à 1 864 document(s)

Choisir le nombre de documents : 1864 / 1864 Tout

Choisir les documents classés :
 Par pertinence & qualité Par pertinence Aléatoirement

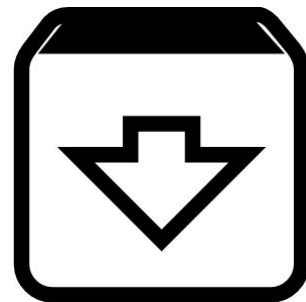
Échantillon de résultats

Über den Niedergang der Kompositionskunst: eine Technisch-Kritische Unters...	Meyer's Anvil: Revisiting the Schema Concept Vasil Byros : HEINRICH SCHENKER ; WILLIAM DRABKIN : Music Analysis 2005	Sonata Form and the Problem of Second-Theme Beginnings Mark Richards : Music Analysis 2013
---	---	--

Bouton "Télécharger" directement connecté à ISTEIX-DL

3.

ISTEX-DL



ISTEX-DL : Rappel

- Téléchargement de masse des documents ISTEX
 - ⇒ Permet de récupérer un jeu de documents à des fins de TDM
- Application web de type "single page"
- Formulaire en 3 étapes
 - (requête, formats, téléchargement)
- Fonctionnalités avancées
 - (historique, partage, récupération, etc.)
- Branchée directement sur l'API ISTEX (<https://api.istex.fr>)
 - ⇒ Résultats identiques aux autres sites ISTEX

ISTEX-DL : accès

<https://dl.istex.fr>

23 millions de documents
littérature scientifique dans tous les domaines
9 307 revues et 348 636 ebooks

Testez ISTEX : indiquez un titre, des mots-clés ou un DOI



**! Authentification nécessaire
pour télécharger des documents en texte intégral**

ISTEX-DL : Rappel

Téléchargez un corpus ISTEX

Vous êtes membre de l'Enseignement supérieur et de la Recherche et vous souhaitez extraire un corpus de documents ISTEX ?
3 étapes suffisent pour récupérer une archive zip sur votre disque dur.

1. Requête

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

Recherche classique | Recherche par ARK

immunology

Nombre de caractères restants : 67 000 ✓

Limite du nombre de documents souhaités : 3000

Choisir les documents classés :
 Par pertinence Aléatoirement

2. Formats et types de fichiers

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

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

3. Télécharger

- *v1 publiée en sept. 2017*
- ...
- *v4.16 en nov. 2018*



ISTEX-DL : 2020, une pluie de nouveautés

- Publication des versions 4.19 (01/2020) et 4.21 (11/2020)
 - 4.19 axée sur la **fiabilisation et l'amélioration des performances**
 - 4.21 axée sur la **nouvelle ergonomie** et l'ajout de **fonctionnalités**
- Billets de blog :
 - *La nouvelle version d'ISTEX-DL repousse les limites*
 - *ISTEX-DL se met sur son 4.21 et sort le grand jeu*

ISTEX-DL : Amélioration des performances

- Meilleure réactivité
lors du clic sur "télécharger"



- Téléchargement jusqu'à 100 000 documents
contre 6 000 auparavant

Équation booléenne ⓘ Identifiants ARK ⓘ Import de fichiers

immunology


L'équation saisie correspond à 947 013 document(s) ⚠

Choisir le nombre de documents ⓘ : 100000 / 100000 Tout

ISTEX-DL : Levée des limitations

- Saisie de requêtes booléennes de grande longueur

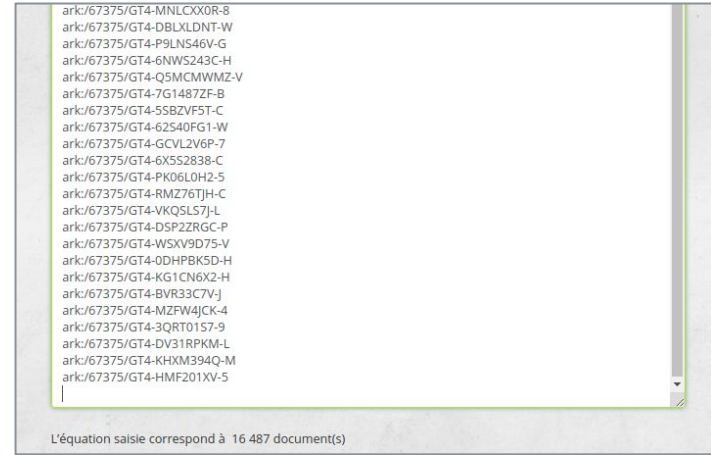
ex. avec 95201 caractères



Équation booléenne ⓘ Identifiants ARK ⓘ Import de fichier ⓘ

phytosociolog* shrubland* vegetal* woodland* amphiphyt* heliophyt* hygrophyt* lithophyt* orophyt* pyrophyt* sciaphyt* sclerophyt* xerophyt*) NOT ((/fungi|use?s?)/ NOT (mycorrhiz* ectomycorrhiz* endomycorrhiz*)) (/bacteri(a|um)/ NOT (actinorhiz* rhizobia azorhizob* bradyrhizob* sinorhizob*)/viruse?s?/ protozoa* protist*))) AND language:"eng" AND qualityIndicators.pdfVersion:[1.2 TO *] AND qualityIndicators.score:[3.0 TO *] AND (publicationDate:[1950 TO *] OR copyrightDate:[1950 TO *]) OR (abstract:(species OR genus OR genera) AND (plantae viridiplantae /gymnosperm(s)? /ae)/ "plant biology" "plant genetic" "plant genetics" "plant genetical" "plant genetically" "plant taxonomy" "plant physiology" "plant morphology" "plant reproduction" "plant reproductive" "plant community" "plant communities" "seed plant" "seed plants" "adventitious plant" "adventitious plants" charophyt* chlorophyt* chlorophyc* rhodophyt* ulvophyt* rhodobiont* glaucophyt* stonewort* ("green alga" OR "green algae" OR "green algal") NOT (cyanophyt* cyanobacter* xanthophyt*)) "red alga" "red algae" "red algal" ((macroalga* seaweed*) NOT ("brown alga" "brown algae" "brown algal" "phaeophyt* fucophyt* chlosophyt*)) "albicidal plant" "albicidal

L'équation saisie correspond à 89 518 document(s)



```
ark:/67375/GT4-MNLCX0R-8
ark:/67375/GT4-DBLXLDNT-W
ark:/67375/GT4-P9LNS46V-G
ark:/67375/GT4-6NWS243C-H
ark:/67375/GT4-Q5MCMWMZ-V
ark:/67375/GT4-7G148ZF-B
ark:/67375/GT4-5SBZVF5T-C
ark:/67375/GT4-62540FG1-W
ark:/67375/GT4-GCVLZ6P-7
ark:/67375/GT4-6X552838-C
ark:/67375/GT4-PK0610H2-5
ark:/67375/GT4-RMZ76TJH-C
ark:/67375/GT4-VKQSL57J-L
ark:/67375/GT4-DSP2ZRG-C
ark:/67375/GT4-WSXV9D75-V
ark:/67375/GT4-0DHPBK5D-H
ark:/67375/GT4-KG1CN6X2-H
ark:/67375/GT4-BVR33C7V-J
ark:/67375/GT4-MZPW4JCK-4
ark:/67375/GT4-3QRT0157-9
ark:/67375/GT4-DV31RPKM-L
ark:/67375/GT4-KHXM394Q-M
ark:/67375/GT4-HMF201XV-5
```

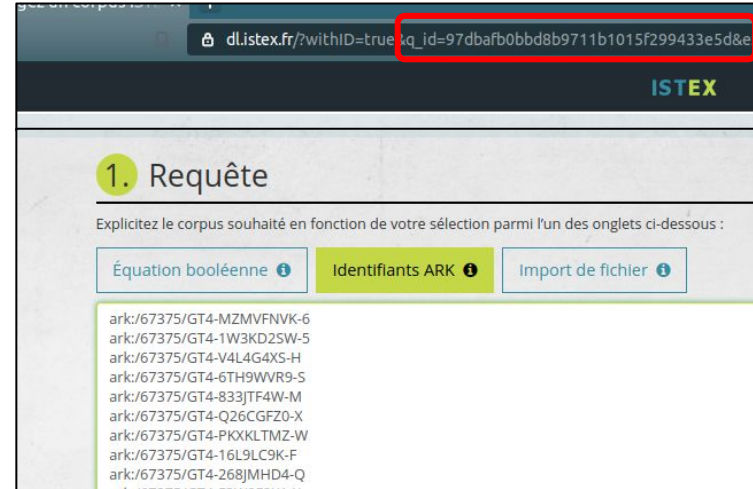
L'équation saisie correspond à 16 487 document(s)

ex. avec 16487 identifiants

- Utilisation de longues listes d'identifiants ARK

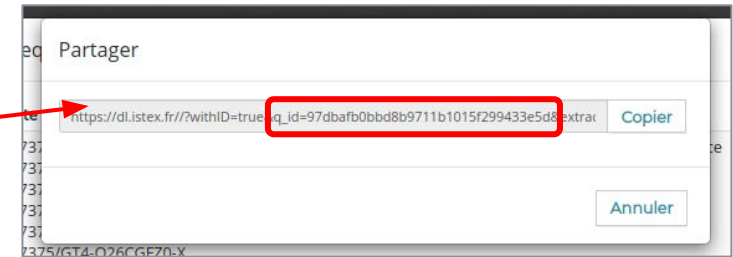
ISTEX-DL : Levée des limitations

- Des identifiants de requêtes longues (`q_id`) pour l'historique et le partage



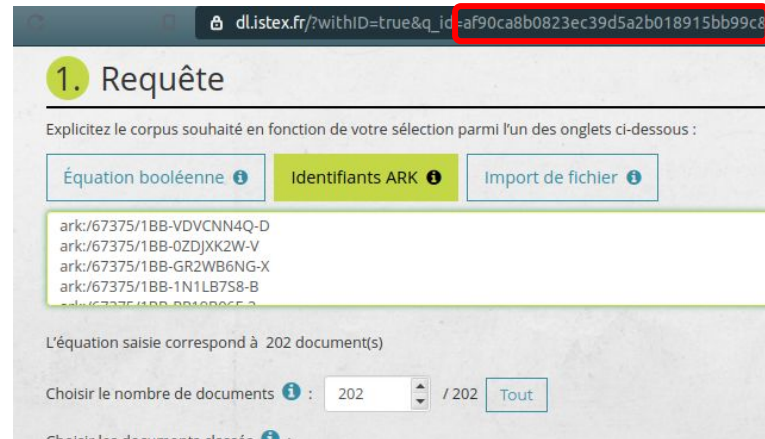
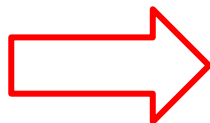
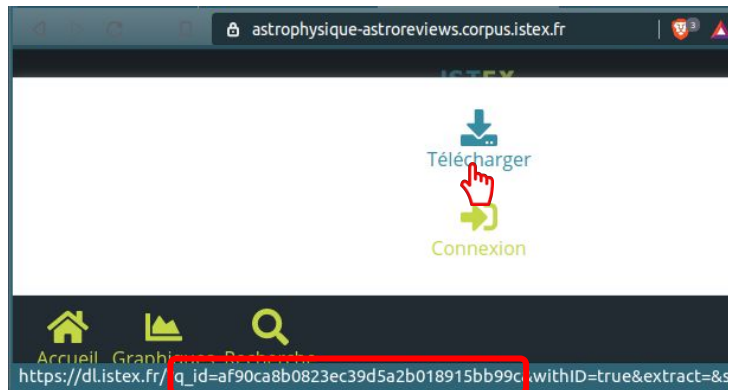
A screenshot of the "Historique des requêtes" dialog box. It displays a table with columns for "#", "Date", "Requête", "Formats", "Nb. docs", "Tri", and "Actions". The "Requête" column contains long ARK identifiers. A red box highlights the share icon in the "Actions" column of the first row.

#	Date	Requête	Formats	Nb. docs	Tri	Actions
1	Tue, 05 Jan 2021 09:59:39 GMT	ark:/67375/GT4-MZMVFNVK-6 ark:/67375/GT4-1W3KD2SW-5 ark:/67375/GT4-V4L4G4XS-H ark:/67375/GT4-6TH9WVR9-S ark:/67375/GT4-833JTF4W-M ark:/67375/GT4-Q26CGFZ0-X ark:/67375/GT4-PKXKLTMZ-W ark:/67375/GT4-16L...	metadata[json]	16 487	qualityOverRelevance	



Exemple d'utilisation des q_id

- Téléchargement d'un corpus depuis data.istex.fr



ISTEX-DL : Une ergonomie repensée

Téléchargez un corpus ISTEX

Vous êtes membre de l'Enseignement supérieur et de la Recherche et vous souhaitez extraire un corpus de documents ISTEX ?
3 étapes suffisent pour récupérer une archive zip sur votre disque dur.

1. Requête

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

Recherche classique | Recherche par ARK

brain AND language:fre

Nombre de caractères restants : 67 000 ✓

Limite du nombre de documents souhaités : 3000

Choisir les documents classés :
 Par pertinence Aléatoirement

2. Formats et types de fichiers

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

<input type="checkbox"/> Texte Intégral	<input type="checkbox"/> Métadonnées	<input type="checkbox"/> Enrichissements
<input type="checkbox"/> PDF	<input type="checkbox"/> XML	<input type="checkbox"/> multicat
<input type="checkbox"/> TEI	<input type="checkbox"/> MODS	<input type="checkbox"/> nb
<input type="checkbox"/> TXT	<input type="checkbox"/> Annexes	<input type="checkbox"/> refBibs
<input type="checkbox"/> OCR	<input type="checkbox"/> Couvertures	<input type="checkbox"/> teeft
<input type="checkbox"/> ZIP		<input type="checkbox"/> unitex
<input type="checkbox"/> TIFF		

3. Télécharger

- Exemples
- Réinitialiser
- Récupérer
- Partager
- Historique

1. Requête

Explicitez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous :

Équation booléenne | Identifiants ARK | Import de fichier | Exemples

brain AND language:fre

Choisir le nombre de documents : 0 / 0 Tout

Choisir les documents classés :
 Par pertinence & qualité Par pertinence Aléatoirement

2. Usage

Cliquez sur l'usage visé pour votre corpus :

DOC TDM	TDM
Usage personnalisé	Lodex
	Analyse graphique / Exploration de corpus
CHOISIR CET USAGE	USAGE SÉLECTIONNÉ

3. Téléchargement

Niveau de compression : Compression moyenne

Format de l'archive : ZIP TAR.GZ

Télécharger

Réinitialiser | Récupérer | Partager | Historique

ISTEX-DL : Une ergonomie repensée

3 Modes de sélection des documents :

- Équation booléenne
- Liste d'identifiants ARK
- Fichiers .corpus

1. Requête

Explicitez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous :

① **Équation booléenne** ⓘ ② Identifiants ARK ⓘ ③ Import de fichier ⓘ

brain AND language:fre

ISTEX-DL : Une ergonomie repensée

① Équation booléenne :

Explicitez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous :

Équation booléenne ⓘ Identifiants ARK ⓘ Import de fichier ⓘ

host.title:immunology AND title:leucocyt* AND publicationDate:[2000 TO *] NOT Immunotherapy

L'équation saisie correspond à 18 document(s)

Choisir le nombre de documents ⓘ : / 18

ISTEX-DL : Une ergonomie repensée

② Liste d'identifiants ARK :

Explicitez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous

Équation booléenne ⓘ Identifiants ARK ⓘ Import de fichier ⓘ

```
ark:/67375/WNG-DBP6SNPT-4
ark:/67375/WNG-JT95B6NR-C
ark:/67375/WNG-F1FPFFF8-D
ark:/67375/6H6-PZZ39TFK-Z
ark:/67375/6H6-127N2MWG-8
ark:/67375/6H6-Q4V0RPNN-X
ark:/67375/6H6-TOC2QM55-X
ark:/67375/6H6-LOZVX6S9-6
ark:/67375/6H6-1HZQPHMT-P
ark:/67375/6H6-06PV859X-c
```

L'équation saisie correspond à 10 document(s)

Choisir le nombre de documents ⓘ : / 10

ISTEX-DL : Une ergonomie repensée

③ Import de fichier .corpus

Explicitez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous :

Équation booléenne ⓘ Identifiants ARK ⓘ Import de fichier ⓘ

Sélectionnez votre fichier

```
#
# Fichier .corpus
#
query      : host.title:immunology AND
title:leucocyt* AND publicationDate:[2000 TO *] NOT
Immunotherapy
date       : 2021-1-4
total      : 10
[ISTEX]
ark ark:/67375/WNG-DBP6SNPT-4
ark ark:/67375/WNG-JT95B6NR-C
ark ark:/67375/WNG-F1PFFFF8-D
ark ark:/67375/6H6-PZZ39TFK-Z
ark ark:/67375/6H6-127N2MWG-8
ark ark:/67375/6H6-Q4V0RPNN-X
ark ark:/67375/6H6-T0C2QM55-X
ark ark:/67375/6H6-L0ZVX6S9-6
ark ark:/67375/6H6-1HZQPMMT-P
ark ark:/67375/6H6-06PV859X-C]
```

Explicitez le corpus souhaité en fonction de votre sélection parmi l'un des onglets ci-dessous :

Équation booléenne ⓘ Identifiants ARK ⓘ Import de fichier ⓘ

Import du fichier .corpus terminé

Modifiez en sélectionnant un autre fichier

Fichier file.corpus analysé. 10 identifiants ont été parcourus. (Attention, le nombre des documents disponibles au téléchargement peut être inférieur si certains identifiants ne sont pas trouvés par le moteur de recherche)

L'équation saisie correspond à 10 document(s)

ISTEX-DL : Une ergonomie repensée

- Saisie du nombre de documents facilitée
- Prévisualisation des 6 premiers résultats avec rebond vers le texte intégral

Choisir le nombre de documents : / 100000

Choisir les documents classés :
 Par pertinence & qualité Par pertinence Aléatoirement

Échantillon de résultats

Neurosteroid Analogues. 8. Structure-Activity Studies of N-Acylated 17a-Az... <i>COVEY Douglas F. ; HAN Mingcheng ; KUMAR ...</i> Journal of Medicinal Chemistry 2000	A Platinum Supramolecular Square as an Effective G-Quadruplex Binder and T... <i>KIELTYKA Roxanne ; ENGLEBIENNE Pablo ; FA...</i> Journal of the American Chemical Society 2008	Map of Unve Analytical Proc
Emissive Terbium Probe for Multiphoton in Vitro Cell Imaging <i>LAW Ga-Lai ; WONG Ka-Leung ; MAN Cornelia...</i> Journal of the American Chemical Society 2008	CCD imaging of basal bioluminescence in larval fireflies: clues on the ana... <i>V. R. Viviani ; F. M. Okawachi ; V. Scors...</i> Photochemical & Photobiological Sciences 2008	Stronglypho Meroditerpen Sponge... <i>WARABI Kaoru</i> Journal of Natu

JACS COMMUNICATIONS

A Platinum Supramolecular Square as an Effective G-Quadruplex Binder and Telomerase Inhibitor

Roxanne Kieltyka,¹ Pablo Englebienne,¹ Johans Fakhoury,¹ Chantal Autexier,^{1,2} Nicolas Moïssié,^{1,2} and Hanadi F. Sleiman^{1,1}

Departments of Chemistry, Anatomy and Cell Biology, McGill University, Montreal QC, Canada

Received February 25, 2008; E-mail: hanadi.sleiman@mcgill.ca

In recent years, significant attention has been devoted to the role of G-quadruplexes in cancer. These structures can arise from the folding of the single-stranded G-rich 3'-end of the telomere into a planar tetrad of guanine bases, held together by a Hoogsteen hydrogen-bonded array.¹⁻³ Small molecules that bind and stabilize this structure have been demonstrated to inhibit the enzyme telomerase, which is active in 85–90% of cancer cells and inactive in somatic cells.⁴ In addition, stabilization of G-quadruplexes by small molecules has recently been shown to inhibit the transcriptional activity of some oncogenes.⁵⁻⁷ Thus, the G-quadruplex motif has emerged as a promising target for the design of selective antitumor therapeutics.^{8,9}

Conventional G-quadruplex binders are organic molecules that possess large π -aromatic surfaces and positive charges near the center of the molecule or on side arms to increase affinity to the grooves of the quadruplex.¹⁰⁻¹⁴ These features have been designed considering the G-quadruplex structure, but also what can be

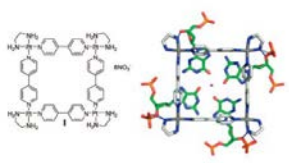


Figure 1. Left: Structure of the platinum molecular square 1. Right: Average structure of the complex between 1 and 22-mer DNA G-quadruplex. Complex 1 is in gray and blue, and hydrogen bonding to phosphates is shown by blue dotted lines.

geometries for effective targeting of G-quadruplexes and for the

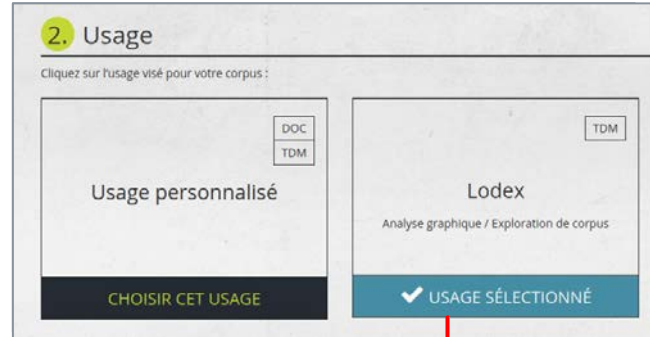
ISTEX-DL : Une ergonomie repensée

- Choix des formats par type d'usage

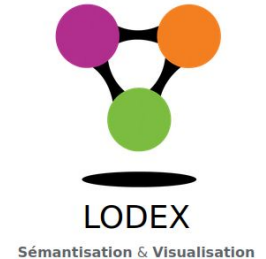


ISTEX-DL : Une ergonomie repensée

- Choix des formats par type d'usage

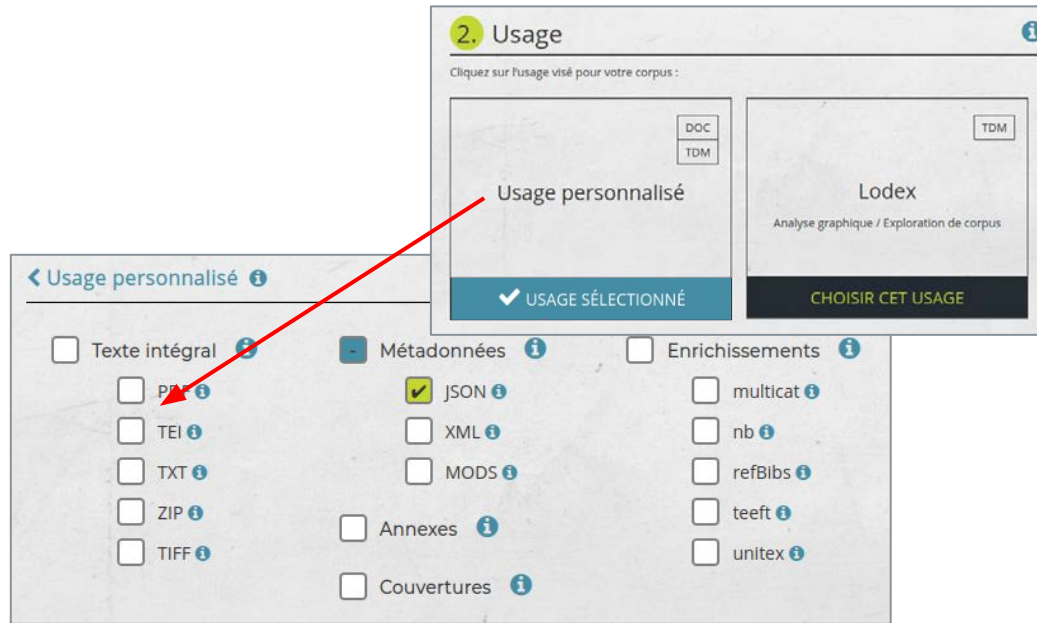


Sélection automatique
du format JSON,
compatible avec le
logiciel LODEX



ISTEX-DL : Une ergonomie repensée

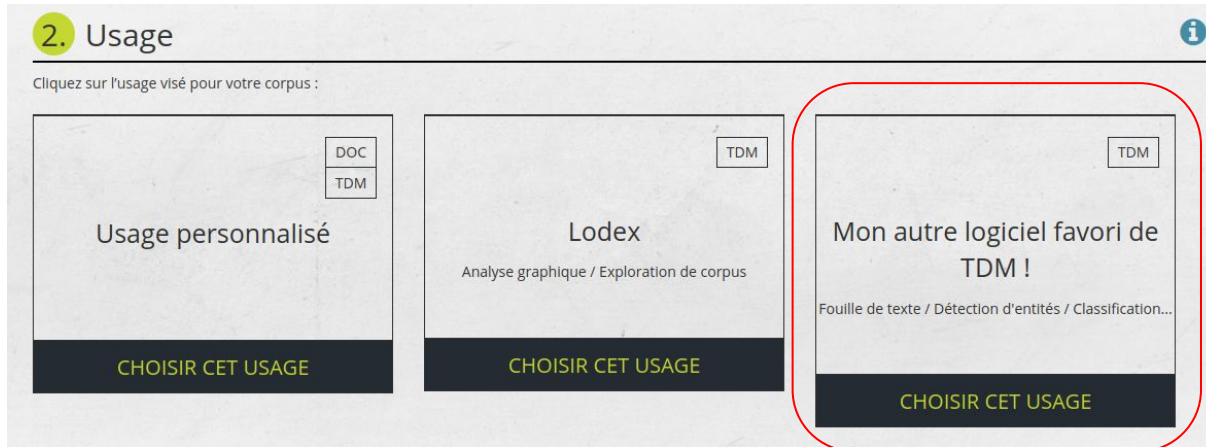
- Choix des formats par type d'usage



Sélection à la carte, en fonction des besoins des utilisateurs

ISTEX-DL : Une ergonomie repensée

- Choix des formats par type d'usage



Coming soon...

ISTEX-DL : Une ergonomie repensée

- Choix du format de l'archive et du taux de compression
(utile selon bande passante et espace disque de l'utilisateur)

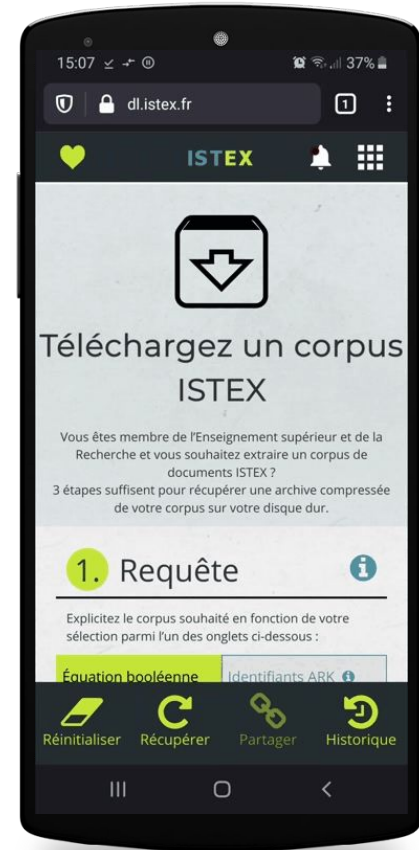
3. Téléchargement

Niveau de compression : Format de l'archive : ZIP TAR.GZ

Télécharger 
Taille estimée > 2 Go

ISTEX-DL : Une ergonomie repensée

- Refonte de l'ergonomie selon une démarche UX
- Interface "responsive", pleinement compatible avec les mobiles



ISTEX-DL : et pour 2021 ?

- De nouveaux usages
 - Formats d'export compatibles Cortext, Gargantext, ou autre outil TDM
 - Ajout d'autres formats d'export (CSV ou autre)
- De nouvelles fonctionnalités
 - Constructeur avancé de requêtes booléennes (ou lien direct avec démonstrateur ISTEX)
 - Ajout de filtres orientés TDM (PDF texte, XML structuré...)
 - ...
- **Vos idées sont les bienvenues !**



Liens utiles

Quelques adresses



Se connecter :

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

S'authentifier :

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

Contact & Informations



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

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

Se tenir informé :

- Blog ISTE_X : <https://blog.istex.fr/>
- Plateforme Twitter : [@ISTEX_Platform](https://twitter.com/ISTEX_Platform)

Merci !

Avez-vous des questions ?
