



Atelier JDEV 2017

Interroger et exploiter l'API ISTEX

JDEV 2017 - Infrastructures logicielles et science ouverte
<http://devlog.cnrs.fr/jdev2017/t5>

Nicolas Thouvenin, Stéphane Gully, Dominique Lechaudel
3 juillet 2017

Objectifs

- Comprendre ce qu'est l'API
 - A quoi sert-elle ?
 - Pour qui ?
- Apprendre à interroger l'API
 - Recherches simples et complexes
 - Accès aux documents sous différents formats
 - Résolveur OpenURL
- Exercices
 - Moissonner l'API avec le harvester
 - IHM au dessus de l'API
 - Programmer l'API avec NodeJS

ISTEX qu'est-ce que c'est ?

- Initiative d'excellence programme « Investissements d'Avenir »
- Accès en ligne aux collections rétrospectives de la littérature scientifique dans toutes les disciplines
- Pour l'ensemble de la communauté de l'enseignement supérieur et de la recherche

ISTEX qu'est-ce que c'est ?

ISTEX

L'excellence documentaire pour tous

“ Construire le socle
de la bibliothèque scientifique
numérique nationale. ”



ANR-10-ID-X-0001-02

WWW.ISTEX.FR

www.licencesnationales.fr



Documentation
et Tutoriels sur les API

api.istex.fr



Site de
démonstration

demo.istex.fr



Widget de
démonstration

widgets.istex.fr



Blog
technique

blog.istex.fr



Listes
de discussion

listes.istex.fr



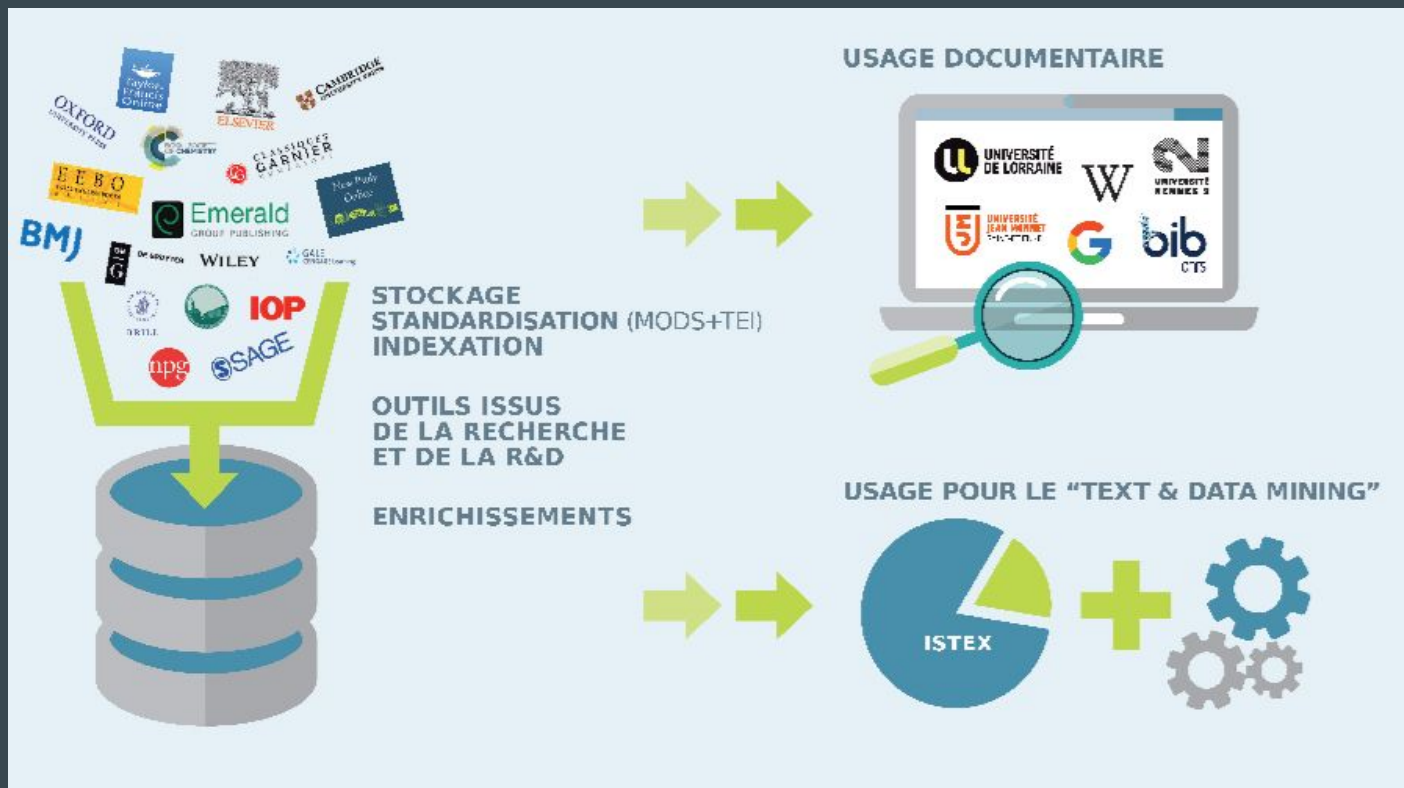
Comptes Twitter

[@Istexdev](https://twitter.com/Istexdev)

[@Projet_ISTEX](https://twitter.com/Projet_ISTEX)

“ ISTE^X plus de **18,5 millions de documents** provenant de **19 éditeurs**
plus de **7 500 titres** et **13 000 ebooks** entre 1406 et 2015. ”

ISTEX qu'est-ce que c'est ?



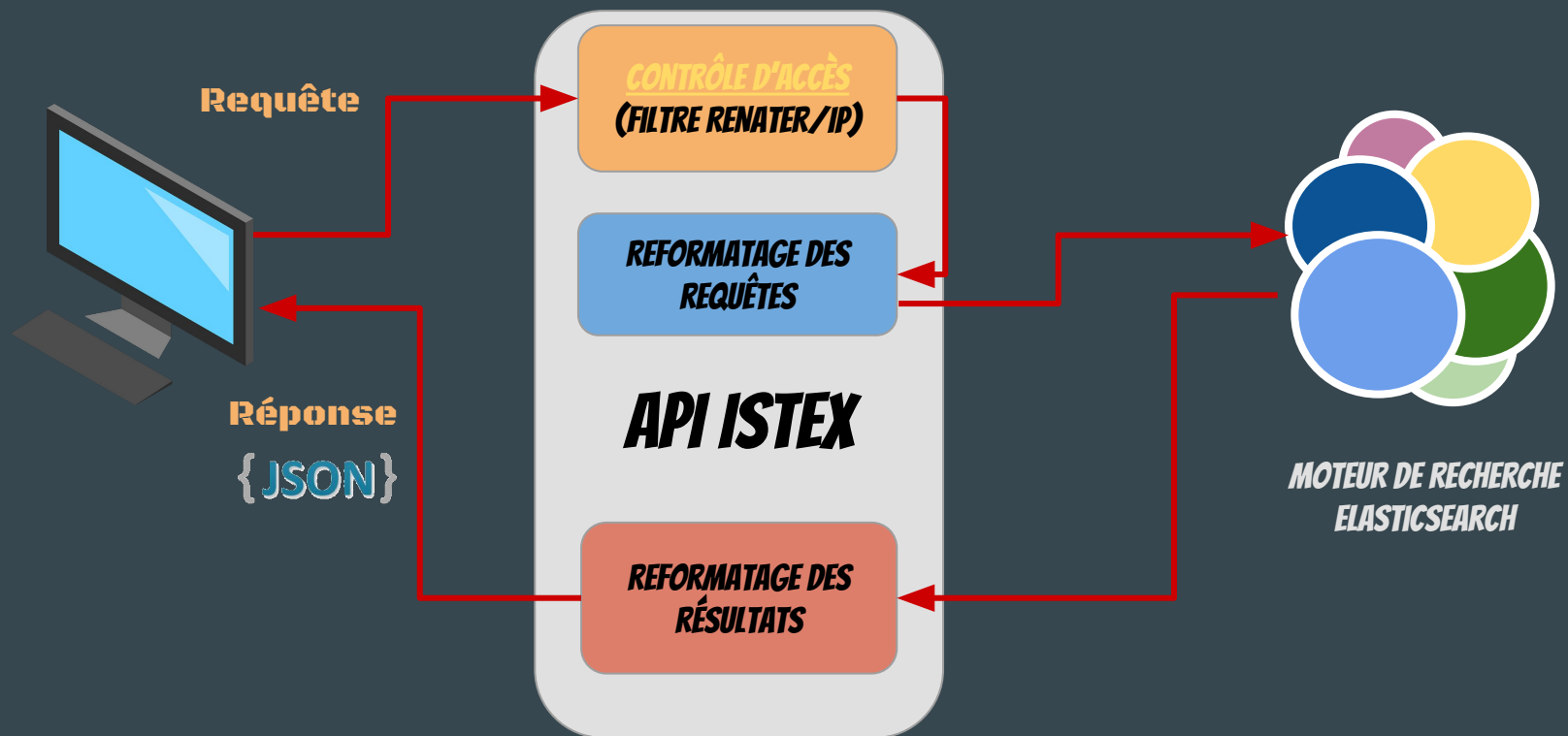
API ISTEK : Les grands principes

- API = **A**pplication **P**rograming **I**nterface
- Pas pour une utilisation humaine directe
- Destinée à être intégrée à une IHM :
 - Site web
 - Application
 - Widget
 - etc.

API ISTEK : Les grands principes

- Et pour ISTEK alors ?
 - Les outils de base fournis concernent la recherche et la lecture des ressources
 - API de type “RESTful”
 - Met en oeuvre des principes standards pour les API orientés ressources
 - Les requêtes à l’API utilisent
 - Le protocole standard HTTP pour le “transport”
 - La syntaxe de requêtes Lucene pour la formulation des recherches
 - Le format JSON pour la structuration des réponses
 - Les utilisateurs directs sont les webmasters/développeurs
 - Manipulent l’API via une surcouche (script, widget, IHM, bibliothèque...)
 - Les utilisateurs finaux sont les membres de l’ESR
 - Vont utiliser la surcouche proposée par les webmasters ou un script maison

API ISTEK : Le trajet de votre requête



Théorie

Usage basique : avant toute chose

- La route pour lire la documentation :
 - <https://api.istex.fr/documentation>
 - Redirection vers la documentation sur <https://api.istex.fr>
- Accès à l'API par :
 - <https://api.istex.fr>
- La route pour vérifier son authentification :
 - <https://api.istex.fr/auth>

Usage basique : avant toute chose

- Routes complémentaires
 - Accès à la liste des corpus : <https://api.istex.fr/corpus>
 - Connaître la version de l'API : <https://api.istex.fr/properties>
 - Le démonstrateur : <http://demo.istex.fr>
 - Les widgets: <https://widgets.istex.fr>
 - Le dépôt Github ISTEEX : <https://github.com/istex>
 - Le démonstrateur des consultations : <http://bibliomap.istex.fr>

Usage basique : la recherche

- Recherche simple

- <https://api.istex.fr/document/?q={votre requête}>
 - Ex : <https://api.istex.fr/document/?q=comics>

- Choix du champ avec le caractère ‘:’

- Ex : [host.eissn.raw:"1476-4687"](#)
- Liste des champs possibles : <https://api.istex.fr/documentation/fields>
- Tous sont interrogés par défaut, fulltext compris
- Tous sont disponibles en 2 versions :
 - version analysée :
 - insensible à la casse lors de la recherche
 - découpage en mots selon séparateurs habituels (dont le tiret, mais pas l’apostrophe)
 - typage spécifique (textuel, entier ou date) selon le type de données
 - version brute, non analysée, utile pour les recherches “exactes”

Usage basique : la recherche

- Syntaxe Lucene de base
 - Opérateurs “OR” (par défaut) et “AND” (ou “+”)
 - Ex : q=comics AND host.eissn.raw:1476-4687
 - Exclusion de terme avec ‘-’ (à privilégier par rapport à “NOT”)
 - Ex : q="death star" -"star wars"
 - Wildcards ? et *
 - Ex : q=title:lang*ge AND language:fre
 - Ex : q=title:*stellar
 - Ex : q=title:(thermal AND polari?ation)

Usage basique : la recherche

- Syntaxe avancée de Lucene

- Expressions régulières

`title:/[frc]at/`

- Importance du parenthésage pour les regroupements

`(Luke OR Leia) AND Skywalker`

`title:(+plastic -bertrand)`

- Recherche par intervalle

`qualityIndicators.abstractCharCount:[2000 TO *]`

`title.raw:{ape TO appliance}`

Usage basique : la recherche

- Syntaxe avancée de Lucene

- “Term boosting”

`“star wars”^2 skywalker`

- Recherche “fuzzy”

`?q=title:france~`

- Recherche de “proximité”

`?q=title:"pomme terre"~1`

`?q=title:"return jedi"~2`

Exercices



Usage basique : la recherche

- Petits exercices pour s'entraîner :

1. On souhaite chercher des résultats possédant le terme gravitation

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

2. On veut maintenant afficher que les documents écrits par Einstein

<https://api.istex.fr/document/?q=gravitation>
[AND author.name:Einstein](#)

3. Comme précédemment, mais en limitant la date de publication entre 1910 et 1920

<https://api.istex.fr/document/?q=gravitation>
[AND author.name:Einstein](#)
[AND publicationDate:\[1910 TO 1920\]](#)

Usage basique : la recherche

- Petits exercices pour s'entraîner :

4. On souhaite cette fois être plus précis en acceptant le terme gravitation ou les termes relativity et Relativität, étant donné qu'Einstein était allemand

```
https://api.istex.fr/document/?q=\(gravitation OR relativity\*\)  
AND author.name:Einstein  
AND publicationDate:\[1910 TO 1920\]
```

5. Pour finir, on veut afficher également les articles écrits par Max Planck, et les prioriser par rapport à ceux d'Einstein

```
https://api.istex.fr/document/?q=\(gravitation OR relativity\*\)  
AND author.name:\(Einstein OR \"Max Planck\"^2\)  
AND publicationDate:\[1910 TO 1920\]
```

Théorie

Usage basique : l'affichage des résultats

- C'est sympa tout ça, mais je ne vois que des ids et des titres...
 - Les champs renvoyés par défaut par l'API sont l'id ISTEEX, le titre et le score du document
 - Pour afficher plus de champs :
 - `https://api.istex.fr/document/?q={votre requête}&output={champs souhaités}`
 - Les champs demandés sont séparés par des virgules
 - L'id est toujours renvoyé, mais utiliser l'option n'affiche plus le titre
 - On peut afficher tous les sous-champs d'un champ "field" avec : **field.***
 - Pour afficher tous les champs renseignés dans le document :
 - `https://api.istex.fr/document/?q={votre requête}&output=*`
- On teste ?
 - On souhaite afficher en plus du titre et de l'id, le nom de l'auteur
 - Encore mieux, on veut tout savoir de l'auteur et pas uniquement son nom !



Usage basique : l'affichage des résultats

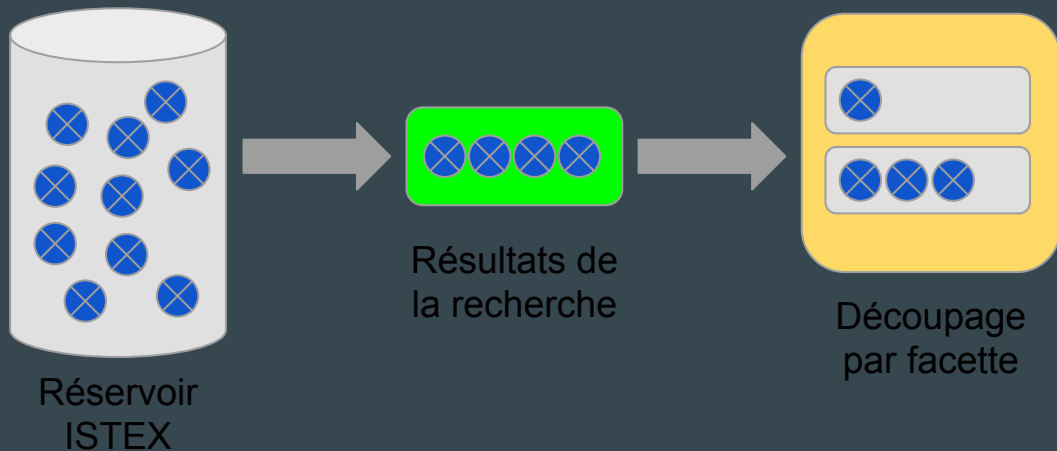
- C'est déjà mieux, mais on peut faire encore plus...
 - Pour sélectionner le nombre de résultats à afficher sur la page :
 - `https://api.istex.fr/document/?q={votre requête}&size={taille}`
 - Le maximum de résultats à afficher ne peut excéder 5000
 - Pour pouvoir paginer les résultats :
 - `https://api.istex.fr/document/?q={votre requête}&from={numéro}`
 - Quatre champs supplémentaires donnent des liens directs :
 - `firstPageURI` : la première page
 - `lastPageURI` : la dernière page
 - `nextPageURI` : la page suivante
 - `prevPageURI` : la page précédente
- On teste ?
 - On souhaite maintenant afficher tous les résultats sur la même page

Usage basique : utilisation des facettes

- Euh... Quésako ?

- Les facettes sont des moyens de filtrer une collection de données en choisissant un ou plusieurs critères. Les facettes sont complémentaires de la requête, et se basent sur celle-ci.

```
api.istex.fr/document/?q=einstein&facet=corpusName
```



Usage basique : utilisation des facettes

- Euh... Quésako ?

- Pour afficher des facettes :
 - `https://api.istex.fr/document/?q={votre requête}&facet={champs}`
 - Tous les champs interrogeables sont "facettables", sauf quelques exceptions :
 - ex de champs interdits : doi, title, facet
 - Les champs demandés sont séparés par des virgules
 - Le type de facette et ses options dépendent du type du champ interrogé
(<https://api.istex.fr/documentation/facets/#facettes-dynamiques-selon-le-mapping>)

- On teste ?

- On souhaite avoir la [liste de tous les auteurs](#) présents dans les résultats
- On souhaite en plus savoir [toutes les années de publication](#) présentes

Les différents formats de données

- **Méta-données** : <https://api.istex.fr/document/4DDACF7F59E85A50A7BFA53896DA6891AFDC466B/metadata>
 - XML, le format éditeur. Dans de rares cas, l'éditeur nous a fourni du MRC (MARC)
 - MODS
- **Plein-texte** : <https://api.istex.fr/document/4DDACF7F59E85A50A7BFA53896DA6891AFDC466B/fulltext>
 - PDF, TXT (si PDF texte)
 - ZIP (contenant l'ensemble de la ressource physique)
 - TEI, conjointement avec les méta-données
- **Annexes** : <https://api.istex.fr/document/4DDACF7F59E85A50A7BFA53896DA6891AFDC466B/annexes>
 - Des formats très variés : JPG, MOV, MP4, GIF, etc...
- **Couvertures** : <https://api.istex.fr/document/4DDACF7F59E85A50A7BFA53896DA6891AFDC466B/covers>
 - Formats images : JPG, GIF, etc...
- **Enrichissements** : <https://api.istex.fr/document/4DDACF7F59E85A50A7BFA53896DA6891AFDC466B/enrichments>
 - TEI (et TEI consolidé)

Authentications dans l'API

Login/mdp dédié pour les JDEV :
istex@inist.fr / jnFRa5QXP7hW7ix

- 4 modes d'authentification:

- **Fédération d'identités** :
 - Authentification Direct via RENATER (**auth = fede**)
 - Token (**auth = jwt**)
- **Adresse IP** (**auth = ip**)
- **Ldap** (**auth = ldap**)

- Demande d'authentification grâce à “**auth=ip,jwt,fede,ldap**”

- Par exemple:
 - <https://api.istex.fr/document/07B79D329D26C4FF72C647E76C68C3BBF9532C94/fulltext/pdf?auth=fede>
- **Cas Interdits** :
 - <https://api.istex.fr/auth?auth=fede,ldap> ou <https://api.istex.fr/auth?auth=ldap,fede>
 - Pas plus de 3 modes d'authentification à demander en même temps :
<https://api.istex.fr/auth?auth=fede,ip,jwt,ip,fede>

L'usage du résolveur OpenURL

Standardisé (Z39.88-2004), le résolveur permet d'accéder directement au plein-texte via un DOI, ou via des méta-données suffisamment exhaustives.

Exemples :

- https://api.istex.fr/document/openurl?rft_id=info:doi/10.1136/acupmed-2012-010150
- <https://api.istex.fr/document/openurl?rft.au=Richard%20James%20Clark&rft.issn=0964-5284&rft.volume=30&rft.issue=4&rft.spage=298>

Avec option “noredirect” :

- <https://api.istex.fr/document/openurl?rft.au=Richard%20James%20Clark&rft.issn=0964-5284&rft.volume=30&rft.issue=4&rft.spage=298&noredirect>

Depuis [Google Scholar](#), [Outils de Découverte](#) et extension ISTEEX <https://addons.istex.fr>

- “Rosetta: The ESA comet rendezvous mission” Acta astronautica Elsevier

Travaux Pratiques

Travaux pratiques

- Exercice 1 : IHM au dessus de l'API ISTEEX avec les widgets
- Exercice 2 : moissonner ISTEEX en ligne de commande
- Exercice 3 : programmer l'API ISTEEX avec NodeJS



Exercices : pré-requis

Login/mdp dédié pour les JDEV :
istex@inist.fr / jnFRa5QXP7hW7ix

Installation de NodeJS / NPM

```
mkdir ~/jdev2017 && cd ~/jdev2017
wget https://nodejs.org/dist/v8.1.2/node-v8.1.2-linux-x64.tar.xz
tar xf ./node-v8.1.2-linux-x64.tar.xz
PATH=$HOME/jdev2017/node-v8.1.2-linux-x64/bin:$PATH

node -v      # v8.1.2
npm -v       # 5.0.3
```

Exercices : pré-requis

Installation de NodeJS / NPM Windows

1. Installation de GIT pour windows :
<https://git-for-windows.github.io/>
2. Installation de NodeJS pour Windows :
<https://nodejs.org/en/download/>

3. Lancement d'un terminal :

Démarrer / Toutes les Applications / Git Bash

4. Test :

```
cd /d
npm install -g echo-cli
echo "Hello ISTEK"
```



Exercice 1 : IHM au dessus de l'API ISTEEX avec les widgets

Création d'un mini site web proposant des IHM de recherche des ressources documentaires hébergées dans la plateforme ISTEEX.

The screenshot displays the ISTEEX search interface. At the top, there is a search bar containing the text 'brain' and a 'Rechercher' button. Below the search bar, there is a link for 'Recherche avancée'. The search results section shows 'Environ 1 551 227 résultats (0.72 secondes)'. A pagination bar shows numbers 1 through 10, with a right arrow. A dropdown menu is set to 'Pertinence'. The first search result is titled 'A most eloquent and panegyric petition to the Prime Minister; by Harriet Air-Brain. Dedicated to Peter Pindar, Esq.'. Below the title, it says 'Pas de résumé' and 'Harriet Air-Brain.'. The snippet reads: 'A most eloquent and panegyric petition to the Prime Minister; by Harriet Air-Brain. Dedicated to Peter Pindar, Esq. vol., page -'. There are three yellow tags: 'Type : Other', 'Ecco', and 'Publié en'. At the bottom left, there are links for 'Texte complet' and 'Métadonnées', with icons for document, PDF, XML, and NCRS. On the right side, there is a section 'Affiner votre recherche' with a 'Corpus' dropdown menu. The corpus list includes: Elsevier (675 680), Wiley (426 442), Springer (146 018), Bmj (109 220), Oup (89 036), Sage (55 318), and Nature (14 264).

Exercice 1 : IHM au dessus de l'API ISTEK avec les widgets

Installation du serveur web (http-server)

```
mkdir -p ~/jdev2017/widgets/  
cd ~/jdev2017/widgets/  
npm install -g http-server
```

```
http-server
```

<http://127.0.0.1:8080>

Index of /

Node.js v8.1.2/ [ecstatic](#) server running @ 127.0.0.1:8080

Exercice 1 : IHM au dessus de l'API ISTEX avec les widgets

Création d'une page HTML vierge (basée sur bootstrap)

```
touch ~/jdev2017/widgets/index.html
```

```
1 <!DOCTYPE html>
2 <html lang="fr">
3   <head>
4     <meta charset="utf-8" />
5     <title>ISTEX</title>
6     <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css" />
7   </head>
8   <body>
9     <h1>Hello, world!</h1>
10
11     <script src="https://ajax.googleapis.com/ajax/libs/jquery/1.12.4/jquery.min.js"></script>
12     <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/js/bootstrap.min.js"></script>
13   </body>
14 </html>
```

<https://github.com/touv/istex-jdev2017/blob/master/exercice1/index1.html>

Exercice 1 : IHM au dessus de l'API ISTEX avec les widgets

Ajout d'un début de mise en page : d'un logo, d'un layout bootstrap et d'un peu de texte dans la page HTML.



ISTEX

About

Contact

Ressources documentaires ISTEX

Collections rétrospectives de la littérature scientifique dans toutes les disciplines de la recherche, proposées grâce à la politique d'acquisition de ressources documentaires à l'échelon national initiée par le projet [ISTEX](#). L'accès au texte intégral est réservé aux établissements publics français.



Exercice 1 : IHM au dessus de l'API ISTEEX avec les widgets

Ajout du widget ISTEEX de recherche dans la page HTML
<https://github.com/istex/istex-widgets-angular>



Ressources documentaires ISTEEX

Collections rétrospectives de la littérature scientifique dans toutes les disciplines de la recherche, proposées grâce à la politique d'acquisition de ressources documentaires à l'échelon national initiée par le projet ISTEEX. L'accès au texte intégral est réservé aux établissements publics français.

Votre requête ici ... 

[Recherche avancée](#)

```
<div class="container">
  <h1>Ressources documentaires ISTEEX</h1>
  <p class="lead">Collections rétrospectives de la littérature
  l'échelon national initiée par le projet <a href="http://w
  <istex-search></istex-search>
</div>
```

<https://github.com/touv/istex-jdev2017/blob/master/exercice1/index3.html>

Exercice 1 : IHM au dessus de l'API ISTEMX avec les widgets

Ajout du widget ISTEMX de liste des résultats et les facettes dans la page HTML.

The screenshot shows the ISTEMX search results page. At the top, there is a navigation bar with 'JDEV2017', 'ISTEX', 'About', and 'Contact'. The main heading is 'Ressources documentaires ISTEMX', followed by a descriptive paragraph. Below this is a search bar with the text 'Votre requête ici ...' and a 'Rechercher' button. The search results show 'Recherche avancée' with 'Environ 18 216 851 résultats (2,84 secondes)'. A pagination bar shows '1 2 3 4 5 6 7 8 9 10 >'. The first result is 'Clinical Pharmacy and Hospital Drug Management' by C. F. George, from 'Postgraduate Medical Journal vol.58, page 803 -'. Below the result, there are three tabs: 'Texte complet', 'Métadonnées', and 'Enrichissements'. At the bottom right, there are filters for 'Type : Book-reviews Bmj' and 'Publié en 1982'.

```
<div class="container">
  <h1>Ressources documentaires ISTEMX</h1>
  <p class="lead">Collections rétrospectives de l'
  échelon national initiée par le projet <a href
  </a>
  </p>
  <istex-search></istex-search>
  <istex-results></istex-results>
  <istex-facets></istex-facets>
</div>
```

The screenshot shows the facet menu on the right side of the search results page. It is titled 'Corpus' and lists various categories with their respective counts. The categories include: Elsevier (45,000), Wiley (45,000), Springer (45,000), Bmj (45,000),oup (45,000), Sage (45,000), Nature (45,000), Diagoniter-journals (45,000), Ing (45,000), Elsevier (45,000), and Qualité de fulltext (45,000). Below the list is a 'Corpus' dropdown menu and a search bar.

<https://github.com/touv/istex-jdev2017/blob/master/exercice1/index4.html>

Exercice 1 : IHM au dessus de l'API ISTEEX avec les widgets

Positionnement des widgets dans la page HTML.

Ressources documentaires ISTEEX

Collections rétrospectives de la littérature scientifique dans toutes les disciplines de la recherche, proposées grâce à la politique d'acquisition de ressources documentaires à l'échelon national initiée par le projet ISTEEX. L'accès au texte intégral est réservé aux établissements publics français.

Recherche avancée

Environ 1 551 227 résultats (1.67 secondes)

Pertinence

Affiner votre recherche

Corpus

- Elsevier 675 680
- Wiley 426 442
- Springer 146 018
- Bmj 109 220
- Oup 89 036
- Sage 55 318
- Nature 14 264
- Degruyter-journals 10 557
- Iop 7 650
- Emerald 6 086

Qualité du fulltext

0 10

Genre

- Research-article 898 070
- Article 367 163

Texte complet

Métadonnées

Remarkable case of a boy, who lost a considerable portion of brain, and is recovered, without detriment to any faculty, mental or corporeal. By R...
R. Leny.

```
<div class="container">
  <div class="row">
    <h1>Ressources documentaires ISTEEX</h1>
    <p class="lead">Collections rétrospectives de la littérature scientifi
l'échelon national initiée par le projet <a href="http://www.istex.fr">
  </div>
  <div class="row">
    <div class="col-xs-12 col-md-8"><istex-results></istex-results></div>
    <div class="col-xs-6 col-md-4"><istex-facets></istex-facets></div>
  </div>
</div>
```

<https://github.com/touv/istex-jdev2017/blob/master/exercice1/index5.html>

Exercice 1 : IHM au dessus de l'API ISTEK avec les widgets

- C'est terminé !

Exercice 2 : moissonner ISTEK en ligne de commande

ISTEK propose une API REST directement interrogeable avec cURL. Nous allons lancer quelques commandes avec cURL pour l'illustrer.

Pré-requis : installer jq pour faciliter la lecture des JSON en ligne de commande

```
cd ~/jdev2017/node-v8.1.2-linux-x64/bin
wget https://github.com/stedolan/jq/releases/download/jq-1.5/jq-linux64
mv jq-linux64 jq && chmod +x ./jq
jq --version # jq-1.5
```



Exercice 2 : moissonner ISTEK en ligne de commande

Chercher avec cURL les 5 premiers documents contenant le mot clé “comics”

```
curl "https://api.istex.fr/document/?q=comics&size=5" | jq .
```

```
{
  "total": 7383,
  "nextPageURI": "https://api.istex.fr/document/?q=comics&size=5",
  "firstPageURI": "https://api.istex.fr/document/?q=comics&size=5",
  "lastPageURI": "https://api.istex.fr/document/?q=comics&size=5",
  "hits": [
    {
      "title": "Image of science in cartoons and its relationship with science",
      "id": "42CEC75BF1851E3383864913AD4F20BC220E980C",
      "score": 20.161226
    },
    {
      "title": "Magazines and journals about comic books",
      "id": "A5F6806D8CAA4F3A801ABA1B0E7BA5AC2D4C4456",
      "score": 19.813581
    },
    {
      "title": "A brief glossary of comic book terminology",
      "id": "4B0BE7CF37A0A988C0B99E1F715A6578651BDFC3",
      "score": 19.743092
    },
    {
      "title": "A librarian's guide to independent comics: Part 1",
      "id": "BE468DFACCC789EF12F470303837B2EE432A069B",
      "score": 19.715754
    },
    {
      "title": "Comic books and graphic novels for libraries: What to look for",
      "id": "796F49613E5304494F4D1A413B06323210B6E06C",
      "score": 19.652575
    }
  ]
}
```



Exercice 2 : moissonner ISTEK en ligne de commande

Ne récupérer que le titre des 5 premiers documents grâce à jq

```
curl "https://api.istex.fr/document/?q=comics&size=5" | jq '.hits[].title'
```

```
"Image of science in cartoons and its relationship with the image in comics"  
"Magazines and journals about comic books"  
"A brief glossary of comic book terminology"  
"A librarian's guide to independent comics: Part one, publisher profiles"  
"Comic books and graphic novels for libraries: What to buy"
```

Exercice 2 : moissonner ISTEEX en ligne de commande

Récupérer les PDF des 5 premières réponses

```
ISTEXIDS=$(curl "https://api.istex.fr/document/?q=comics&size=5&output=id" | jq -r  
' .hits[].id')  
for ISTEEXID in $ISTEXIDS  
do  
  echo "Downloading ${ISTEEXID}.pdf"  
  curl -u istex@inist.fr:jnFRa5QXP7hW7ix  
"https://api.istex.fr/document/${ISTEEXID}/fulltext/pdf" > ${ISTEEXID}.pdf  
done
```



```
4B0BE7CF37A0A9 42CEC75BF1851E3 796F49613E53044 A5F6806D8CAA4F BE46BDFACCC789  
88C0B99E1F715A6 383864913AD4F20 94F4D1A413BD63 3A801ABA1B0E7B EF12F470303837B  
578851BDFC3.pdf BC220E980C.pdf 23210B6E06C.pdf A5AC2D4C4456. 2EE432A069B.pdf  
pdf
```



Exercice 2 : moissonner ISTE_X en ligne de commande

ISTEX propose un outil en ligne de commande pour extraire des documents de la plateforme ISTE_X : `istex-api-harvester`.

Installation du script :

```
npm install -g istex-api-harvester
```

Exercice 2 : moissonner ISTEK en ligne de commande

On va faire la même chose : récupérer les 5 premiers PDF mais avec le harvester !

```
istex-api-harvester --username istex@inist.fr \  
--password jnFRa5QXP7hW7ix \  
-q comics --size 5 \  
--fulltext "pdf" --metadata ""
```



1C0A2534A3C2C3
E11327FCDF9204F
753D222AAB0.
fulltext.original.
[pdf](#)



43AEBCD3521909
E2FD56F96272D1
BAD9B5AD4FF7.
fulltext.original.
[pdf](#)



135C6418FD48841
7D59D2C9460F52
6EE3FCE04F1.
fulltext.original.
[pdf](#)



B069949C13A225
C493BFF59F4FCE
ADA488569EB2.
fulltext.original.
[pdf](#)



F28FCE731F0F87B
BF9743FB0273284
252B81CE47.
fulltext.original.
[pdf](#)

Exercice 3 : programmer l'API ISTEEX avec NodeJS

Création d'une mini chaîne de traitement sous forme d'une commande linux, prenant une requête sur l'entrée standard pour afficher les titres et les résumés correspondants :

```
echo "comics" | ./istex
```

Exercice 3 : programmer l'API ISTEEX avec NodeJS

Préparation

```
mkdir jdev2017
cd jdev2017
npm init
npm i -P ezs ezs-basics ezs-istex lodash.pick
touch istex
chmod 755 ./istex
```

Exercice 3 : programmer l'API ISTEK avec NodeJS

1. Initialisation

```
1  #!/usr/bin/env node
2
3  const ezs = require('ezs');
4  const basics = require('ezs-basics');
5  const istex = require('ezs-istex');
6
7  ezs.use(basics);
8  ezs.use(istex);
9
10 process.stdin.resume();
11 process.stdin.setEncoding('utf8');
12 process.stdin
13   .pipe(process.stdout)
14   ;
15
```

```
echo "comics" | ./istex
```

Exercice 3 : programmer l'API ISTEEX avec NodeJS

2. Récupérer l'entrée standard

```
10 function str2qry(input, feed) {
11   if (input) {
12     feed.send({ q: input.trim() });
13   } else {
14     feed.send(input);
15   }
16 };
17
18 process.stdin.resume();
19 process.stdin.setEncoding('utf8');
20 process.stdin
21   .pipe(ezs('stringify'))
22   .pipe(ezs(str2qry))
23   .pipe(ezs('jsonify'))
24   .pipe(process.stdout)
25 ;
```

```
echo "comics" | ./istex
```


Exercice 3 : programmer l'API ISTEEX avec NodeJS

3. Configurer la requête vers ISTEEX

```
18  const output = 'abstract,title';
19
20  process.stdin.resume();
21  process.stdin.setEncoding('utf8');
22  process.stdin
23    .pipe(ezs('stringify'))
24    .pipe(ezs(str2qry))
25    .pipe(ezs('ISTEXQuery', { params: { output, }, }))
26    .pipe(ezs('jsonify'))
27    .pipe(process.stdout)
```

```
echo "comics" | ./istex
```

Exercice 3 : programmer l'API ISTEEX avec NodeJS

4. Récupérer les réponses de l'API ISTEEX

```
20 const pick = require('lodash.pick');
21 function extract(input, feed) {
22   if (input) {
23     feed.send(pick(input.ISTEEX, ['title', 'abstract']));
24   } else {
25     feed.send(input);
26   }
27 };
28
29 process.stdin.resume();
30 process.stdin.setEncoding('utf8');
31 process.stdin
32   .pipe(ezs('stringify'))
33   .pipe(ezs(str2qry))
34   .pipe(ezs('ISTEXQuery', { params: { output, }, }))
35   .pipe(ezs('ISTEXRequest'))
36   .pipe(ezs('ISTEXHits'))
37   .pipe(ezs(extract))
38   .pipe(ezs('jsonify'))
39   .pipe(process.stdout)
```

```
echo "comics" | ./istex
```

Exercice 3 : programmer l'API ISTEK avec NodeJS

5. **Tout** récupérer

```
20 const pick = require('lodash.pick');
21 function extract(input, feed) {
22   if (input) {
23     feed.send(pick(input.ISTEK, ['title', 'abstract']));
24   } else {
25     feed.send(input);
26   }
27 };
28
29 process.stdin.resume();
30 process.stdin.setEncoding('utf8');
31 process.stdin
32   .pipe(ezs('stringify'))
33   .pipe(ezs(str2qry))
34   .pipe(ezs('ISTEKQuery', { params: { output, }, }))
35   .pipe(ezs('ISTEKScroll'))
36   .pipe(ezs('ISTEKHits'))
37   .pipe(ezs(extract))
38   .pipe(ezs('jsonify'))
39   .pipe(process.stdout)
40 ;
```

```
echo "comics" | ./istek
```

Aller plus loin

API demo extension widget status blog listes

ISTEX

L'excellence documentaire pour tous

« Construire le socle de la bibliothèque scientifique numérique nationale. »

Présentation Tests et enquêtes Les ressources La plateforme Editeurs

Vous êtes ici : Accueil

ISTEX : initiative d'excellence en Information scientifique et technique

Le projet ISTEX a pour principal objectif d'offrir, à l'ensemble de la communauté de l'enseignement supérieur et de la recherche, un accès en ligne aux collections rétrospectives de la littérature scientifique dans toutes les disciplines en engageant une politique nationale d'acquisition massive de documentation : archives de revues, bases de données, corpus de textes...

Une volonté nationale

Le projet ISTEX (Initiative d'excellence de l'Information Scientifique et Technique) s'inscrit dans le programme « **Investissements d'Avenir** », initié par le Ministère de l'Enseignement Supérieur et de la Recherche (MESR), dont l'ambition est de **renforcer la recherche et l'enseignement supérieur français sur la scène mondiale**.

Ce projet novateur est porté par **quatre partenaires** : le Centre National de la Recherche Scientifique (CNRS), l'Agence Bibliographique de l'Enseignement Supérieur (ABES), le Consortium Unifié des Etablissements Universitaires et de Recherche pour l'Accès aux Publications Numériques (Couperin.org), et l'Université de Lorraine agissant pour le compte de la Conférence des Présidents d'Université (CPU).

Signée le 19 avril 2012 par l'ANR (Agence Nationale pour la Recherche), l'Etat et le CNRS, la convention d'attribution d'aide au projet s'élève à **60 millions d'euros sur trois ans [1]** dont 54 millions d'euros consacrés aux acquisitions des ressources documentaires et 6 millions d'euros alloués à la création de la plateforme destinée à héberger les données.

Rechercher

Actualités

27 juin 2017
Fermeture temporaire du site internet.

2 juin 2017
Séminaire technique : les 6 et 7 juin 2017 – Présentations

29 mai 2017
Numérique Premium : une nouvelle licence nationale !

» Actualités

Usage basique : avant toute chose

Site de démonstration

Widget de démonstration

Comptes twitter



Documentation et tutoriels sur les API et leur prise en main

Blog technique

Listes de discussion



api-users@listes.istex.fr



@Projet_ISTEX et @istexdev



<http://www.istex.fr>
<https://api.istex.fr>
<http://blog.istex.fr>

Outils développés en internes

- Extensions ISTEEX en version Beta
 - <https://addons.istex.fr>
 - Version pour [Firefox](#)
 - Version pour [Google Chrome](#)
- Extraction de sous-corpus
 - Via la fonctionnalité d'export Zip de l'API ([documentation](#))
 - [https://api.istex.fr/document/?q=troll&extract=fulltext\[pdf\];metadata\[mods\]&size=90](https://api.istex.fr/document/?q=troll&extract=fulltext[pdf];metadata[mods]&size=90)
 - Via le [harvester ISTEEX](#) à disposition sur github

À vous de jouer...

