

# TRAITER LES DONNEES DE LA PRESSE

---

# Constitution du corpus

- Objectif: constituer un corpus exhaustif selon un certain nombre de critères
  - Une équation de recherche bien définie
  - Une sélection raisonnée des sources:
    - Presse francophone/ presse anglophone/ presse française/ presse nationale...
    - Restriction éventuelle aux articles longs si beaucoup d'articles
    - Autre critère de sélection possible sous réserve d'être pertinent
- A noter: les équations de recherche testées
  - Pour ne pas refaire x fois la même chose et pouvoir expliquer les choix effectués
- Enregistrement sous 2 formats:
  - Sous forme pdf (mieux pour l'impression éventuelle)
  - Sous forme html: pour traitement automatisé éventuel

# Une fois le corpus Europresse constitué, que faire?

- Si nombre d'articles < 300 -> les lire en ordre chronologique
- Si nombre d'articles > 300 -> repérer les articles les plus longs/ riches et les lire en ordre chronologique.
- Relever:
  - Les **événements** (manifestation, loi, procès, publication etc.) et leur date
  - Les **acteurs** qui interviennent ou sont cités (individus, mais aussi associations, entreprises, syndicats, administrations etc.), leur **statut** dans la controverse (scientifique, politique, militant, etc.), et leur **position** dans la controverse.
  - Les autres **publications** ou **médias** cités et leurs références: autres articles de presse, articles scientifiques, livres, rapports, reportages TV, documentaires, émissions de radio, sites web, blogs etc.

# Gargantext, une aide au traitement

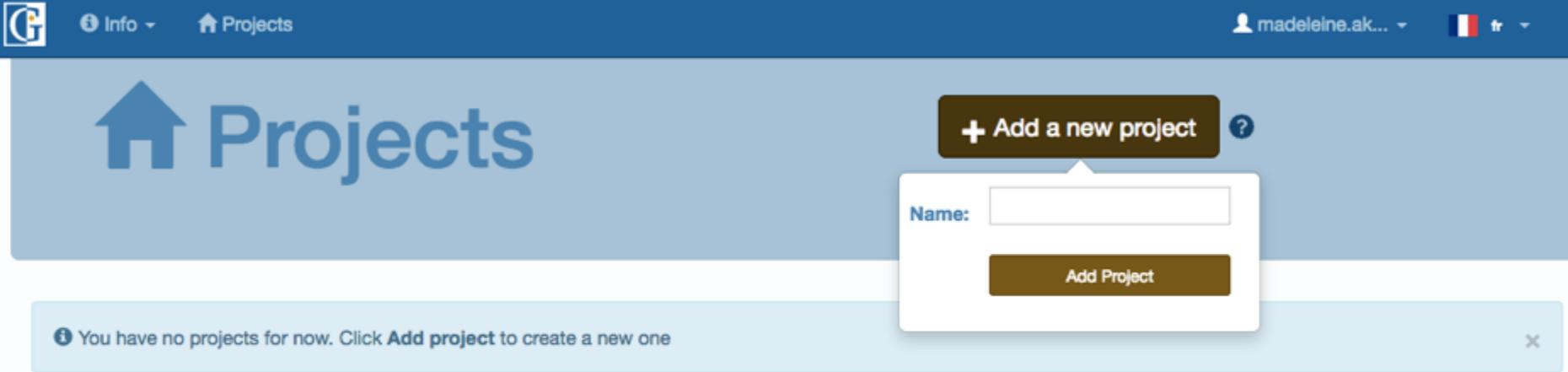


- Plateforme de traitement qui combine:
  - Analyse automatisée du langage
  - Algorithme d'analyse statistique des textes
  - Modules de visualisation
- Adresse: [stable.gargantext.org](http://stable.gargantext.org)
- Utiliser plutôt **Chrome**
- Un login pour chaque groupe

Attention! Gargantext ne fait pas l'analyse à votre place: c'est un outil d'exploration de corpus.

- Il aide à faire une sélection raisonnée des articles à approfondir
- Il doit permettre de récupérer une information pertinente de manière systématique

# L'interface

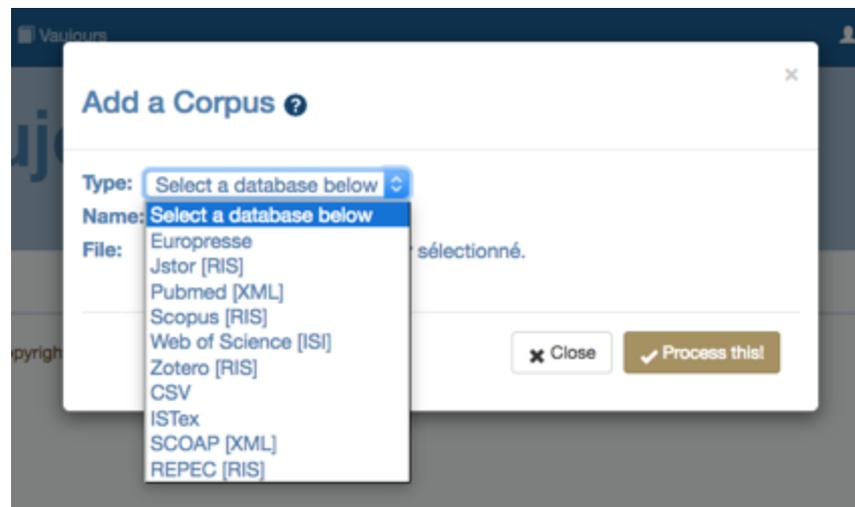


Après le login, une page d'accueil personnalisée sur laquelle on retrouve tous les projets, et qui permet d'ajouter de nouveaux projets.

1<sup>ère</sup> étape: ajouter un **nouveau projet**, puis cliquer sur ce projet. Chaque projet peut comprendre un ensemble de corpus (presse, scientifique, etc.) sur un même thème.

# La constitution des corpus

- Différentes bases de données possibles
- Les fichiers doivent être **zipés** (.zip) avant d'être uploadés
  - Si vous avez plusieurs fichiers Europresse, les ziper ensemble en un seul fichier



- Etape 1: choix du format
- Etape 2: donner un nom au corpus
- Etape 3: choisir le fichier à uploader
- Etape 4: cliquer sur Process
- Etape 5: le nouveau fichier doit apparaître**

## Accueil après téléchargement

Vaujours

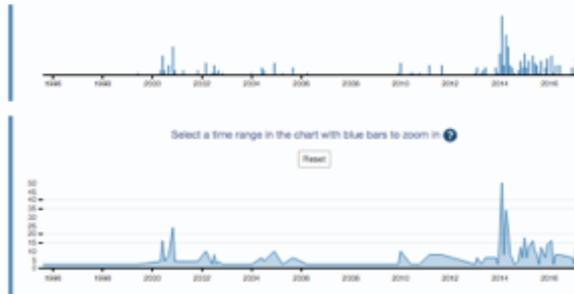
Europresse

EuropresseVaujours

March 1, 2017, 5:06 p.m.

Author(s): madeleine.ak...

Export corpus



Publications by title

Search:   10

Showing 1 to 10 of 263 records Page: Previous 1 2 3 ... 27 Next

Date	Title	Source	★	🗑️
1995/08/09	La Dhuis, l'aqueduc invité chez Mickey Construit au XIXe pour alimenter Paris, il abreuve aujourd'hui Disneyland.	Libération	★	🗑️
1999/06/18	ARMÉE Une nouvelle stratégie foncière	Le Point	★	🗑️
2000/05/09	Courty Le Fort de Vaujours joue son avenir	Le Parisien	★	🗑️
2000/05/30	Courty Le fort de Vaujours inquiète les défenseurs de l'environnement	Le Parisien	★	🗑️
2000/06/05	Courty Les défenseurs de l'environnement déçus	Le Parisien	★	🗑️
2000/06/13	Fort de Vaujours Les Verts demandent la levée du secret militaire	Le Parisien	★	🗑️
2000/06/13	La première bombe atomique française conçue en banlieue de Paris	La Presse	★	🗑️
2000/06/14	POLLUTION. L'ancien centre de recherche du Commissariat à l'énergie atomique de Seine-Saint-Denis bientôt vendu L'encombrant pylône du fort de Vaujours	Le Figaro	★	🗑️

## Accueil après indexation et rechargement de la page

Vaujours

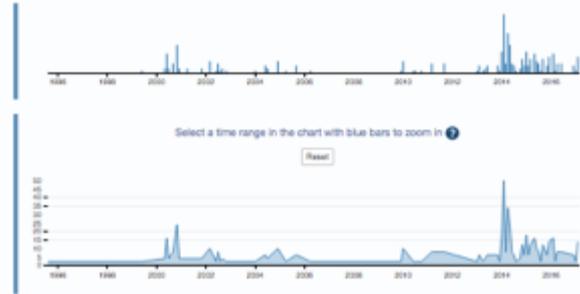
Europresse

EuropresseVaujours

March 1, 2017, 5:06 p.m.

Author(s): madeleine.ak...

Export corpus



Publications by title

Search:   10

Showing 1 to 10 of 263 records Page: Previous 1 2 3 ... 27 Next

Date	Title	Source	★	🗑️
1995/08/09	La Dhuis, l'aqueduc invité chez Mickey Construit au XIXe pour alimenter Paris, il abreuve aujourd'hui Disneyland.	Libération	★	🗑️
1999/06/18	ARMÉE Une nouvelle stratégie foncière	Le Point	★	🗑️
2000/05/09	Courty Le Fort de Vaujours joue son avenir	Le Parisien	★	🗑️
2000/05/30	Courty Le fort de Vaujours inquiète les défenseurs de l'environnement	Le Parisien	★	🗑️
2000/06/05	Courty Les défenseurs de l'environnement déçus	Le Parisien	★	🗑️
2000/06/13	Fort de Vaujours Les Verts demandent la levée du secret militaire	Le Parisien	★	🗑️
2000/06/13	La première bombe atomique française conçue en banlieue de Paris	La Presse	★	🗑️
2000/06/14	POLLUTION. L'ancien centre de recherche du Commissariat à l'énergie atomique de Seine-Saint-Denis bientôt vendu L'encombrant pylône du fort de Vaujours	Le Figaro	★	🗑️

- **Attention:** une fois le corpus uploadé, il faut un certain temps au système pour l'indexer; toutes les fonctions ne sont pas immédiatement disponibles.

Info · Projects · Vaujours · EuropresseVaujours · madeleine.ak... ·

Documents Sources Analytics Graphs Terms

Vaujours

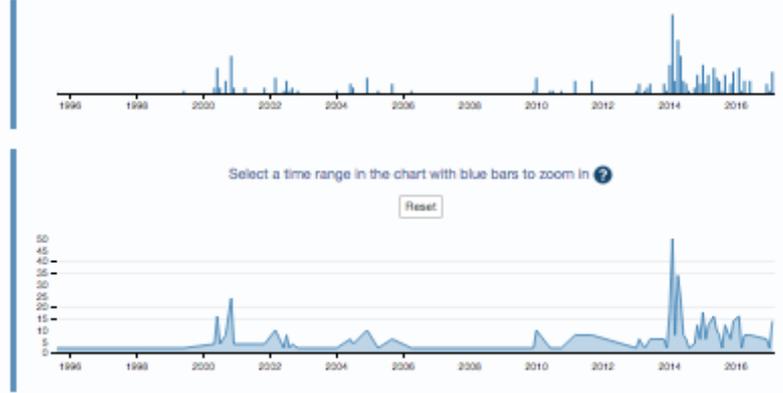
Export corpus ?

Europresse March 1, 2017, 5:06 p.m.

EuropresseVaujours Author(s): madeleine.ak...

Menu

Possibilité d'exporter le corpus en format csv (Excel)



Analyse temporelle et exploration du corpus

Publications by title

Search:  All Show: 10

Showing 1 to 10 of 263 records Pages: Previous 1 2 3 ... 27 Next

Date	Title	Source	☆	All page
1995/08/09	La Dhuis, l'aqueduc invité chez Mickey Construit au XIXe pour alimenter Paris, il abreuve aujourd'hui Disneyland.	Libération	☆	<input type="checkbox"/>
1999/06/18	ARMÉE Une nouvelle stratégie foncière	Le Point	☆	<input type="checkbox"/>
2000/05/09	Courty Le Fort de Vaujours joue son avenir	Le Parisien	☆	<input type="checkbox"/>
2000/05/30	Courty Le fort de Vaujours inquiète les défenseurs de l'environnement	Le Parisien	☆	<input type="checkbox"/>
2000/06/05	Courty Les défenseurs de l'environnement déçus	Le Parisien	☆	<input type="checkbox"/>
2000/06/13	Fort de Vaujours Les Verts demandent la levée du secret militaire	Le Parisien	☆	<input type="checkbox"/>
2000/06/13	La première bombe atomique française conçue en banlieue de Paris	La Presse	☆	<input type="checkbox"/>
2000/06/14	POLLUTION. L'ancien centre de recherche du Commissariat à l'énergie atomique de Seine-Saint-Denis bientôt vendu L'encorement pyralène du fort de Vaujours	Le Figaro	☆	<input type="checkbox"/>

Liste des articles et accès au contenu



# Nettoyer le corpus

Publications by title

Search:  TI AB All Show: 10

Showing 1 to 10 of 263 records Pages: Previous **1** 2 3 ... 27 Next

Date	Title	Source	Star	Trash
1995/08/09	La Dhuis, l'aqueduc invité chez Mickey Construit au XIXe pour alimenter Paris, il abreuve aujourd'hui Disneyland.	Libération	★ <input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
1999/06/18	ARMÉE Une nouvelle stratégie foncière	Le Point	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/05/09	Courtry Le Fort de Vaujours joue son avenir	Le Parisien	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/05/30	Courtry Le fort de Vaujours inquiète les défenseurs de l'environnement	Le Parisien	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/06/05	Courtry Les défenseurs de l'environnement déçus	Le Parisien	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/06/13	Fort de Vaujours Les Verts demandent la levée du secret militaire	Le Parisien	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/06/13	La première bombe atomique française conçue en banlieue de Paris	La Presse	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/06/14	POLLUTION. L'ancien centre de recherche du Commissariat à l'énergie atomique de Seine-Saint-Denis bientôt vendu L'encombrant pyralène du fort de Vaujours	Le Figaro	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/06/14	FORT DE VAUJOURS . Le commissaire enquêteur était un ancien du CEA	Le Parisien	★ <input type="checkbox"/>	<input type="checkbox"/>
2000/06/14	Fort de Vaujours Elus et habitants veulent la vérité	Le Parisien	★ <input type="checkbox"/>	<input type="checkbox"/>

Trash It!

Sélectionner les enregistrements à supprimer

Cliquer sur Trash it!

# Supprimer les doublons

Publications by title

Search:

TI AB

Duplicates (title, date & src)

All  
Favorite documents  
Duplicates (title, date & src)

Show: 10

Showing 1 to 10 of 22 records Pages: Previous 1 2 3 Next

Date	Title	Source	☆	All page
2015/01/23	À la télé Carton pour le fort de Vaujours	Le Parisien	☆	
2015/01/23	À la télé Carton pour le fort de Vaujours	Le Parisien	☆	■
2015/07/16	Comment aider un zadiste qui va en justice	Rue89 (site web)	☆	
2015/07/16	Comment aider un zadiste qui va en justice	Rue89 (site web)	☆	■
2002/03/08	Courty Des décharges sauvages le long du fort de Vaujours	Le Parisien	☆	
2002/03/08	Courty Des décharges sauvages le long du fort de Vaujours	Le Parisien	☆	■
2014/11/05	Courty Le fort de Vaujours dans le "New York Times"	Le Parisien	☆	
2014/11/05	Courty Le fort de Vaujours dans le "New York Times"	Le Parisien	☆	■
2015/07/22	Environnement Fort de Vaujours : l'arrêt des travaux demandé	Le Parisien	☆	
2015/07/22	Environnement Fort de Vaujours : l'arrêt des travaux demandé	Le Parisien	☆	■

# Accès aux articles (Europresse) ou résumés (autres bases de données)

- Sélection d'un **sous-ensemble d'articles** correspondant à une période particulière
- Visualisation des articles en cliquant sur le titre

Vaujours

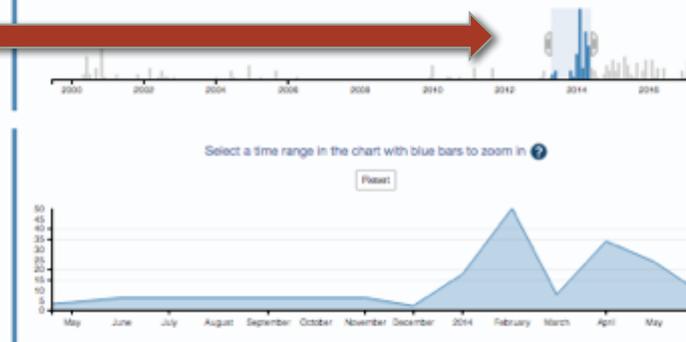
Export corpus

Europresse

EuropresseVaujours

March 1, 2017, 5:06 p.m.

Author(s): madeleine.ak...



Publications by title

Search:

Showing 1 to 10 of 74 records

Pages: Previous 1 2 3 ... 8 Next

Date	Title	Source	★	All pages
2013/05/09	Country Le dernier conseil municipal vote à l'unanimité la révision du PLU Un bois bientôt déclassé pour la future carrière de gypse	Le Parisien	★	□
2013/05/15	CountryLes sauveteurs-déblayeurs s'entraînent sur l'ancien site du CEA Le Fort de Vaujours terrain de jeu des pompiers de Chelles	Le Parisien	★	□
2013/06/13	Country Une association d'amoureux du Fort de Vaujours se bat pour sa sauvegarde. Ils veulent sauver le fort atomique en y créant un musée	Le Parisien	★	□
2013/06/14	clichy-sous-bois Après un an d'installation Les derniers roms du bois de la Couronne évacués	Le Parisien	★	□
2013/06/22	Le Pin-VILLEPARISIS Des visites guidées toute la journée Placo ouvre sa carrière de gypse au public	Le Parisien	★	□
2013/11/01	Un documentaire sur le passé atomique du fort de Vaujours	Le Parisien	★	□

# Les résumés ou articles

- Lecture à l'écran
- En vert, les **termes** qui ont été **indexés** par le système: l'indexation permet de construire des graphes: plutôt pour les corpus d'articles scientifiques

## courtry Le dernier conseil municipal vote à l'unanimité la révision du PLU Un bois bientôt déclassé pour la future carrière de gypse

Le Parisien

source

05/09/13

date

### abstract

Un bois bientôt déclassé pour la future carrière de gypse

"La société plâtrière sollicite la collectivité pour qu'elle fasse évoluer son PLU (plan local d'urbanisme) de manière à permettre la réalisation de son projet d'exploitation." Cette annonce est extraite de la délibération votée à l'unanimité par le conseil municipal de courtry du 4 avril. Le maire (PS) Jean-luc Pilard et les élus ont accepté de réviser le PLU concernant une parcelle classée en espace naturel boisé, qui jouxte le fort de vaujours où bpb placo a pour projet de développer une carrière de gypse à ciel ouvert.

Propriétaire de 33 ha sur l'ancien site du centre de recherche atomique, "bpb placo devrait déposer une demande d'autorisation d'exploitation en préfecture avant 2015", avait déclaré philippe cortial, directeur des mines et carrières du groupe en février. Située au nord du fort de vaujours, la parcelle concernée par la modification du PLU est un espace boisé classé. "Une grande partie n'est déjà plus un bois mais une zone de pâturage", se défend le maire qui promet une compensation "gagnante" puisque "la zone non boisée sera également compensée sur une autre parcelle de la commune, où seront plantés des arbres".

Un risque sanitaire majeur selon des associations

"Nous modifions le PLU à la demande de placo, sinon, nous ne l'aurions pas fait, cela représente beaucoup de travail pour nos services", ajoute l'édile qui assure "que cela se soldera par une enquête publique d'ici septembre". "Depuis quand des entreprises dictent aux communes leur plan local d'urbanisme? courtry n'a même pas attendu le dépôt officiel de demande d'exploitation et le résultat de l'enquête publique qui doit en découler. C'est antidémocratique et irresponsable", dénonce christophe nedelec du collectif d'associations protectrices de l'environnement "Sauvons la dhuis". Il demande "la mise en place d'une véritable concertation associée à une enquête publique approfondie".

# La liste des journaux

- Accessible en **cliquant sur Sources**
- Histogramme de répartition des articles dans les journaux



The screenshot displays a software interface with a blue header bar containing navigation tabs: 'Documents', 'Sources', 'Analytics', 'Graphs', and 'Terms'. The 'Sources' tab is active, showing a breadcrumb path: 'Vaujourns > Europresse > EuropresseVaujourns'. The user is identified as 'madeleine.ak...' and the session started on 'March 1, 2017, 5:06 p.m.' Below the header, there are two histograms. The top histogram shows a distribution of articles across sources, with a 'Reset' button and a zoom instruction: 'Select a time range in the chart with blue bars to zoom in'. The bottom histogram shows a similar distribution. At the bottom of the interface, a table titled 'Publications by source' is displayed, showing the number of publications for various sources.

Title	No. Publs
Le Parisien	202
Le Parisien.fr	12
L'humanité	6
Aujourd'hui en France (site web)	5
AFP Infos Françaises	5
Le Figaro	3
Journal de l'environnement.com	2
La Lettre A	2
La Gazette des communes (site web)	1
NewsPress (français)	1

# Recherche et sélection des articles

The screenshot shows a search interface for 'Publications by title'. The search bar contains 'Christophe Nedelec'. There are filters for 'TI' and 'AB', and a dropdown menu set to 'All'. The 'Show:' dropdown is set to '10'. Below the search bar, it says 'Showing 5 of 5 records' and 'Pages: Previous 1 Next'. The results are displayed in a table with columns: Date, Title, Source, a star icon, and a trash icon labeled 'All page'.

Date	Title	Source	★	All page
2013/02/21	Courtry. Un collectif d'associations alerte sur le risque radioactif et chimique Placo veut creuser sous l'ancien site atomique	Le Parisien	★	<input type="checkbox"/>
2013/05/09	Courtry Le dernier conseil municipal vote à l'unanimité la révision du PLU Un bois bientôt déclassé pour la future carrière de gypse	Le Parisien	★	<input type="checkbox"/>
2014/04/16	Une radioactivité anormale détectée au fort de Vaujours	Le Parisien	★	<input type="checkbox"/>
2014/11/03	Seeing a New Future for Nuclear Site, After the Toxic Dust Has Settled	The New York Times	★	<input type="checkbox"/>
2014/11/06	Montfermeil Réunion publique sur le fort de Vaujours	Le Parisien	★	<input type="checkbox"/>

- S'effectue uniquement sur les termes indexés (liste sous l'onglet Terms)
- Pour rechercher dans le corps des articles ou des résumés, cocher AB
- Permet par exemple de retrouver tous les articles où une personne est citée ou un concept utilisé etc.

# Autre méthode possible



- Il est possible de télécharger le corpus en cliquant sur download
- On obtient un fichier csv

- Permet toutes sortes de traitement et de représentations graphiques
  - Extraction directe de citations par copié collé
  - Histogramme temporel dans un autre format
  - Recherche dans le texte...

# Fichier csv->Excel

On peut utiliser Open Office

Avec Excel: Importer et Origine du fichier = Unicode (UTF-8); séparateur = tabulateurs

	A	B	C	D	E	F	G
1	title	source	publication_year	publication_month	publication_day	abstract	authors
2	Courtry Le Fort de Vaujours joue son avenir	Le Parisien	2000	5	9	<p>&lt;p&gt;Le Fort de Vaujours joue son avenir&lt;/p&gt; &lt;p&gt;L'AVENIR de l'ancien site du Commissariat à l'énergie atomique (CEA), plus connu sous le nom de Fort de Vaujours , se joue peut-être actuellement. A partir d'aujourd'hui et jusqu'au 10 juin, le préfet fait procéder à une enquête publique en mairie de Courtry. Après, il devrait déclarer la vacance de ce lieu aujourd'hui désaffecté. L'armée et le CEA sont les propriétaires de cet endroit qui s'étend, pour deux tiers, sur Courtry et, pour un tiers, sur Vaujours (Seine-Saint-Denis). Au total, 45 hectares sont répartis entre les deux communes. Pourtant, depuis fin 1997, les 22 500 mètres carrés de bâtiments, qui accueillent alors 800 salariés, sont laissés à l'abandon. Pendant une trentaine d'années, ce site a servi de lieu d'expérimentation pour les explosifs chargés de l'allumage des fusées. Pour certains, des essais étaient faits à l'intérieur du Fort. Aujourd'hui, les différents interlocuteurs, dont le maire de Courtry, Michèle Godon, se posent la question de sa reconversion.&lt;/p&gt; &lt;p&gt;Créer des emplois&lt;/p&gt; &lt;p&gt;Le projet s'orienterait vers une séparation des secteurs. La partie où se situe le Fort (30 ha) serait vendue par adjudication, tandis que la partie sud serait cédée à la commune. "Notre priorité est la reprise économique. Il n'est pas question de créer un nouveau quartier, avec des constructions de logements, mais de favoriser l'installation de nouvelles entreprises. Cette partie de la Seine-et-Marne a, en effet, un grand besoin d'emplois", explique-t-on à la mairie. Si la municipalité prospecte dans tous les secteurs d'activités, l'exploitation et la fabrication du plâtre sont une des éventualités. En effet, les anciens bureaux du CEA sont installés sur une butte gypsifère qui constitue une réserve reconnue d'intérêt général. Avant d'envisager les futures acquisitions, la municipalité se préoccupe d'abord de la dépollution du site. "Lorsque le CEA est parti, il nous a remis un bilan somme toute assez positif. Toutefois, il reste encore des produits à l'intérieur des terres. Nous restons donc extrêmement vigilants", conclut la mairie.&lt;/p&gt;</p>	
3	Courtry Le fort de Vaujours inquiète les défenseurs de l'environnement	Le Parisien	2000	5	30	<p>&lt;p&gt;Le fort de Vaujours inquiète les défenseurs de l'environnement&lt;/p&gt; &lt;p&gt;UNE RÉUNION publique et une prolongation de l'enquête sur l'avenir de l'ancien site du commissariat à l'énergie atomique (CEA) de Courtry. Les associations de défense de l'environnement n'en demandent pas moins. "Cette enquête, qui a démarré le 9 mai, doit se terminer le 10 juin. Mais le dossier n'est pas clair. Et il est surtout très complexe", justifie Daniel de Beckers, président des Amis de la terre de la Dhuis. Deux autres associations sont à ses côtés : l'Adebf (Association pour la défense de l'environnement du bois fleuri) et l'ASMSN (Association seine-et-marnaise pour la sauvegarde de la nature). "Nous avons donc écrit au préfet de Seine-et-Marne pour lui demander la tenue d'une réunion début juin. A cette occasion, les habitants du secteur pourraient rencontrer le commissaire-enquêteur, des personnes qui ont travaillé sur le site et des contradicteurs bien sûr", poursuit Daniel de Beckers. Le fort de Vaujours , qui abritait le CEA, est désaffecté depuis fin 1997. Il s'étend sur Courtry et sur Vaujours, et appartient toujours à l'armée et au CEA. Le site, qui a servi de lieu d'expérimentation pour les explosifs chargés de l'allumage des fusées pendant une trentaine d'années, pourrait en partie (à hauteur des deux tiers) être cédé à la commune de Courtry. "Il faut exiger la dépollution du terrain, avant que la ville ne le récupère", martèle Daniel de Beckers.&lt;/p&gt;</p>	

# 2<sup>ème</sup> niveau: Fonction Analytics

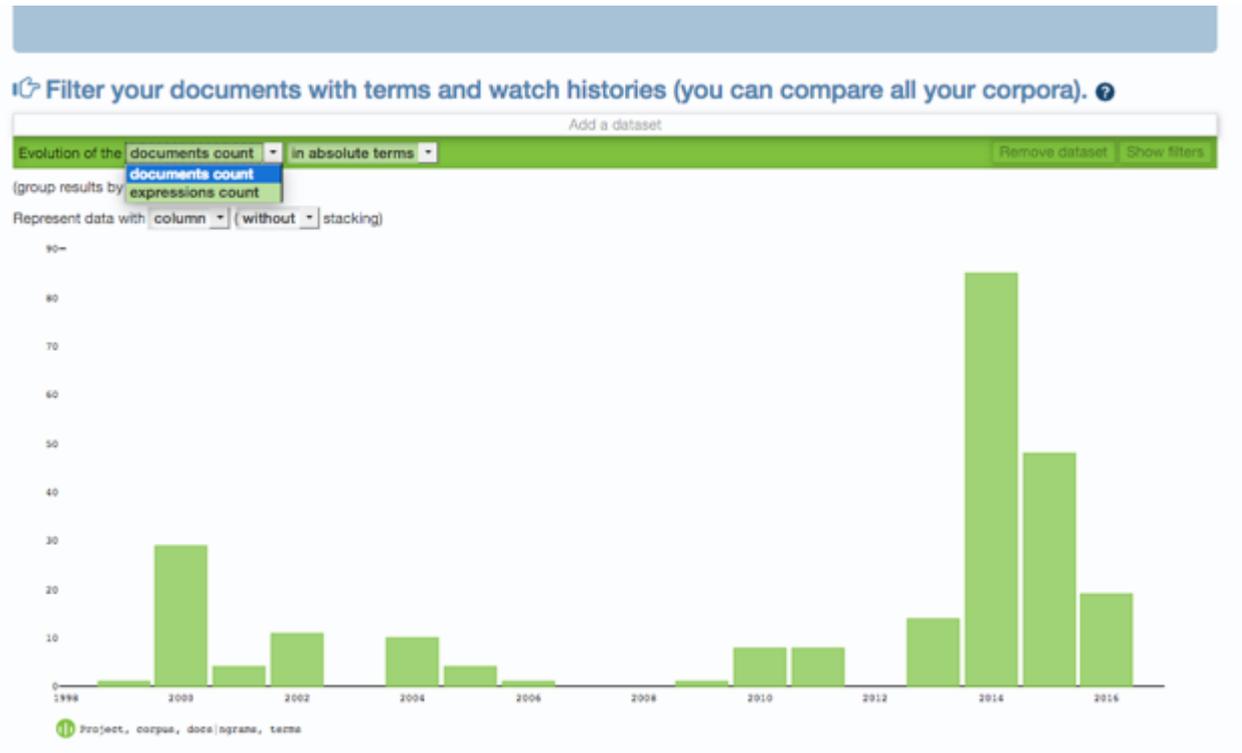
Permet d'analyser et de comparer l'évolution temporelle

- de deux ou plusieurs corpus
- De différents termes ou expressions dans un corpus
- Etc.

Différents formats:

- Échelle de temps
- En valeur absolue/relative
- Histogramme/courbes

Récupération: capture d'écran pour le moment



# Utilisation de filtres

Cliquer sur Show Filters pour faire la recherche

Filter your documents with terms and watch histories (you can compare all your corpora). ?

Add a dataset

Evolution of the **expressions count** in **absolute terms** Remove dataset Hide filters

...restrict to the following corpora: select none select all

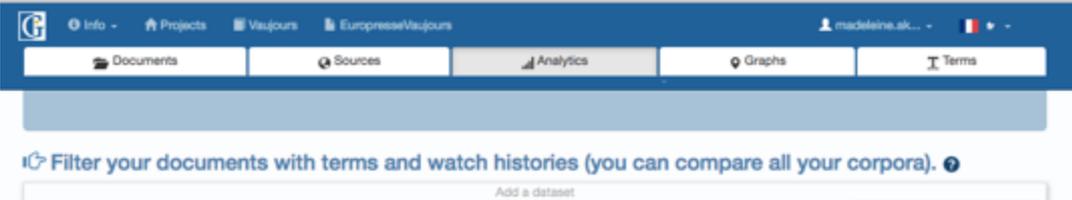
**EuropresseVaujourns**

...only consider documents containing the following expressions:

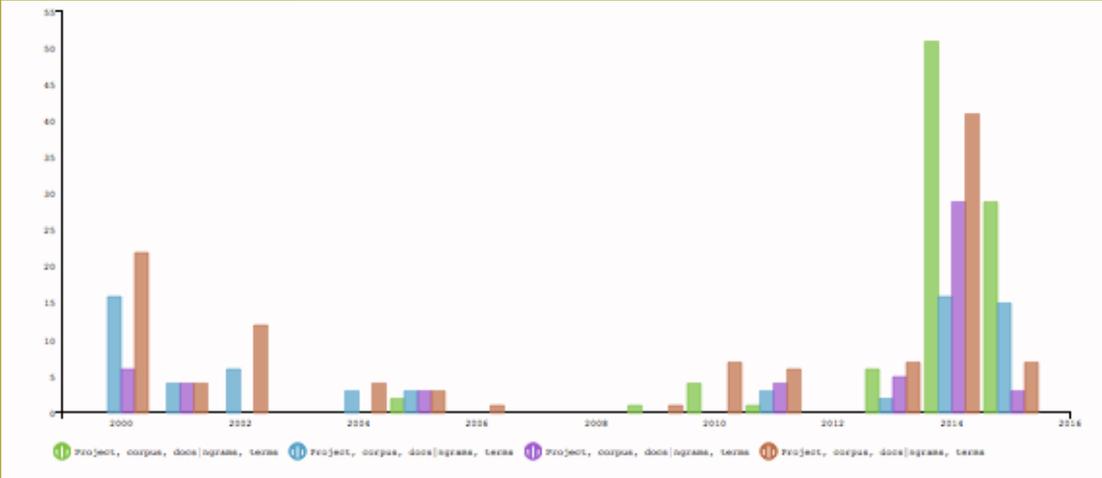
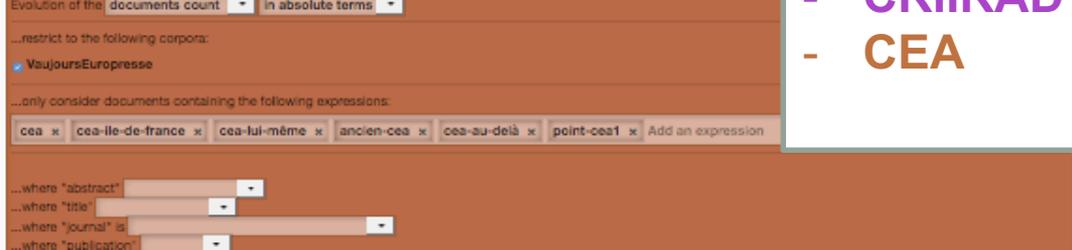
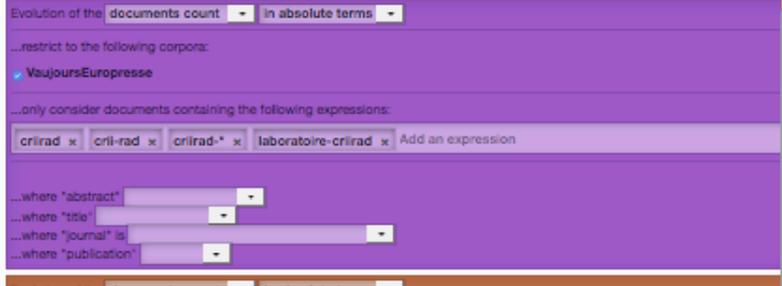
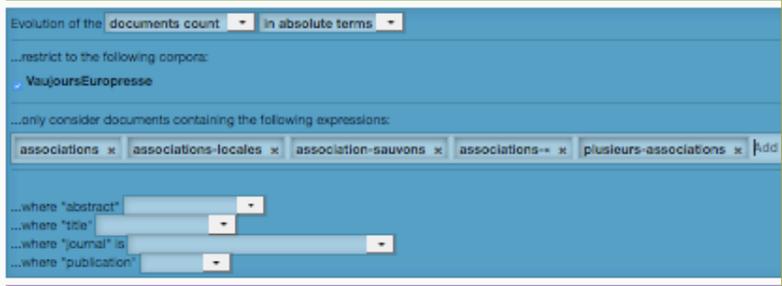
placo|

- placoplatre**
- placo**
- placoplâtre
- bpb **placo**
- société **placoplâtre**
- bp **placo**
- société **placoplatre**
- entreprise **placoplatre**
- usine **placoplatre**
- entreprise **placoplâtre**

Recherche sur les expressions indexées avec suggestions automatiques



Faire Add a Dataset pour rajouter une variable.  
On peut comparer plusieurs corpus à l'intérieur du même projet



Comparaison dans le temps du nombre de publications qui parlent de:

- Placoplâtre
- Associations
- CRIIRAD
- CEA

# En guise de conclusion

- Première étape du traitement: LIRE les articles ou un échantillon raisonné des articles
- Gargantext vous permet dans un second temps:
  - De revenir de manière systématique sur certains aspects de la controverses (acteurs, problèmes, concepts)
  - De vous aider à trouver les réponses à certaines questions que vous vous posez ou certaines hypothèses que vous formulez.
- Gargantext est un outil encore en développement:
  - En cas de difficulté particulière ou de bug, vous pouvez contacter:  
Alexandre Delanoë: [alexandre.delanoe@mines-paristech.fr](mailto:alexandre.delanoe@mines-paristech.fr)