



Depuis 80 ans, nos connaissances  
bâtissent de nouveaux mondes

# INTRODUCTION GÉNÉRALE SUR LES ENJEUX DU *TEXT ET DATA MINING*

PATRICE BELLOT

8 OCTOBRE 2019



cnrs

La Science Ouverte : une révolution nécessaire

8 oct. 2019 Paris (France)

CNRS – INS2I





# 1 QU'EST-CE QUE LE TDM ?

Domaines scientifiques et difficultés

# 1

## QU'EST-CE QUE LA FOUILLE DE DONNÉES / DE TEXTES ?

### Le croisement de plusieurs domaines

- L'analyse de données automatisée
- L'Intelligence Artificielle
- Le Traitement Automatique des Langues

### De la donnée vers l'information vers la connaissance

#### Données ?

Elles peuvent des nombres,  
des signaux, des mots, des  
images...

Elles peuvent être structurées  
ou non, liées ou non...

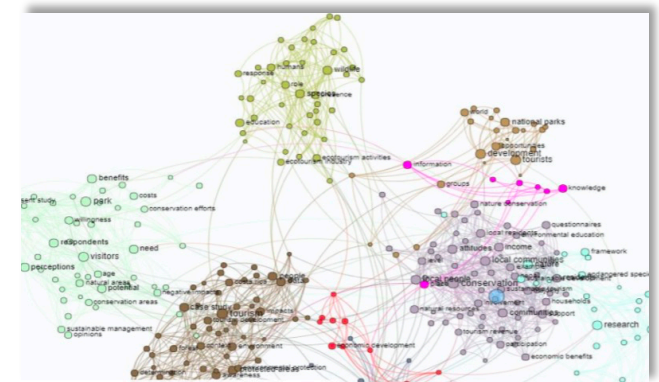
# 2

## DES APPLICATIONS

- Recherche de documents, d'images, de pages Web
- Recommandation automatique de contenus, veille
- Systèmes de questions-réponses
- Résumé automatique
- Cartographie et navigation guidées
- Détection de nouveauté, analyse de sentiments, de tendances...



<https://isidore.science>



Gargantext



# 3

## DES TRAITEMENTS GÉNÉRIQUES

- Extraction de termes, d'entités, d'informations, de relations
- Sélection de descripteurs, filtrage de données
- Calcul de similarités sémantiques
- Classification ou catégorisation automatiques
- Recherche d'information à partir de requêtes
- Structuration et segmentation
- Reconnaissance de formes, de caractères, d'images, de la parole...

# 4

## QUELQUES DIFFICULTÉS...

### Quelle que soit la nature des données :

- Structures peu normalisées, formats variés
- Les fameux V du Big Data : Volume, véracité, variabilité, valeur, vitesse

### Document, texte et langage :

- Données hétérogènes ou multimodales, ambiguës
- Multilinguisme (lexiques, terminologies, syntaxes)

Des difficultés  
génériques.

Des standards  
nécessaires.

Des solutions  
à partager.

### Le Droit et les bonnes pratiques pour la mise en œuvre du TDM :

Briefing  
Requested by the JURI committee



**The Exception for Text and Data Mining (TDM)  
in the Proposed Directive on Copyright in the  
Digital Single Market - Technical Aspects**

*loi française Pour une République Numérique du  
7 octobre 2016 et la directive européenne sur le  
droit d'auteur dans le marché unique numérique  
du 26 mars 2019*

# Trump attaque l'Europe après le feu vert de l'OMC

Les Etats-Unis vont taxer à 10 % les avions et à 25 % une myriade de produits européens

**D**ans la foulée de leur victoire, les Etats-Unis ont annoncé mardi 2 octobre, le droit d'imposer des droits de douane de 10 % sur les avions et de 25 % sur une myriade de produits européens. Un record historique. Le département du Commerce a taxé en priorité les importations en provenance des quatre pays responsables de la dette illégale à Airbus, à savoir la France, le Royaume-Uni, l'Allemagne et l'Espagne. L'administration Trump compte pour l'instant taxer à hauteur de 10 % les avions civils et pénaliser à hauteur de 25 % de nombreuses importations agricoles et industrielles européennes.

Après cette victoire juridique, les Etats-Unis qui demandent la compensation d'une réponse à l'OMC le 14 octobre, qui devrait valider automatiquement leur liste, et prévoient d'appliquer leurs pénalités le 18 octobre. Les Etats-Unis ont remarqué que l'Union européenne (UE) n'a pas le droit d'un plateau de contre-mesures à ces

les décisions par les Etats-Unis sans cesse pressés. Sur la liste des produits taxés à 25 %, publiés par le représentant au commerce Robert Lighthizer figurent les vins français et espagnols, l'huile d'olive espagnole, le whisky écossais, les draps et les couvertures britanniques, le café, les couteaux et les machines allemandes et des fromages de toute l'Europe, secteur très sensible. Me chez les agriculteurs américains frappés par une crise latente. Les produits en cuir n'y figurent plus, ce qui protège la maroquinerie des groupes de luxe français comme LVMH.

C'est une grande victoire pour les Etats-Unis, a déclaré mercredi à la Maison Blanche, Donald Trump, qui a insisté sur la décision de l'OMC concernant une plainte déposée par ses prédecesseurs. Les victoires ont lieu maintenant parce qu'on les

- Sur la liste figurent vins français et espagnols, huile d'olive espagnole, whisky écossais, draps britanniques.**

que je n'aime pas l'OMC et ils vont être stricts que je suis content. Washington, taxe nos produits. Airbus pour plusieurs raisons. D'abord, les compagnies aériennes américaines se nous sommes largement suppos de l'avion leur européen, qui livre chaque année une centaine d'avions aux Etats-Unis. Imposer des droits massifs équivalait à un embargo et serait intenable pour les compagnies américaines. Delta, qui a refusé au American Airlines avaient fait part de leurs inquiétudes ces derniers mois. Les droits de douane seraient imposés sur Delta, à ses employés, et à ses clients, et auraient pas d'incidence sur Airbus, car ces ventes ont déjà été affectées, avait déclaré mercredi la directrice de Delta, Bessing qui avait demandé une taxation à 100 % selon le Wall Street Journal.

C'est pas vraiment en mesure de profiter. Le groupe, né à Seattle, est en pleine reorganisation après le double accident des 737 MAX provoqué par des défauts de conception et qui a coûté au total sept appareils.

Une halte dans le prix. Surtout, les Etats-Unis et les Européens sont jusqu'à présent, évité de surenchérir dans le domaine de l'aéronautique, tant Airbus et Boeing vivant de subventions illégales au regard des règles de l'OMC. Ce serait prêter une visibilité largement transatlantique de (de nombreux rectorats

**L'affaire est loin d'être close, les Etats-Unis ont, eux aussi, été reconnus coupables de subventions illégales à Boeing.**

Avions sont fabriqués aux Etats-Unis. Ce sera se tirer une balle dans le pied, alors que Boeing et Russes subventionnent massivement leur secteur aéronautique et menacent le duopole transatlantique. L'administration Trump évite pour l'instant une guerre commerciale aéronautique que l'irréductible, mais semble donner un coup de pouce opportun à Boeing, pénalisé par le fait fort et le scandale du 737 MAX.

L'affaire est loin d'être close puisque les Etats-Unis ont, eux aussi, été reconnus coupables de subventions illégales à Boeing, l'avionneur américain ayant bénéficié de crédits d'impôt. Depuis 2020, l'OMC doit fixer le montant des sanctions que l'Europe pourra imposer aux importations américaines. « Nous continuons de penser que même si Washington y est autorisé par l'organe de règlement des différends, choisir d'imposer des sanctions maintenant serait contre-productif et de courte vue », a réagi mercredi soir la commissaire européenne au commerce, Cecilia Malmström.

Mercredi 2 octobre, Bruno Le Maire, le ministre français de l'économie et des finances, avait d'ailleurs insisté : « Nous souhaitons faire redescendre les tensions et éviter des sanctions tarifaires qui ne feront que nuire à nos deux économies, au moment où la Chine renforce son industrie aéronautique. [Mais s'ils] choisissent

d'imposer des sanctions, ce serait une erreur économique et politique. Nous serons prêts avec nos partenaires européens à y répondre de façon ferme. »

Les mesures américaines sont censées faire plier les Européens, qui n'ont jusqu'à présent pas flanché face à Donald Trump dans leurs négociations commerciales, ou en tout cas les amener à la table des négociations. La querelle aéronautique, qui a toujours été spécifique, arrive par un hasard du calendrier sur le devant de la scène, alors que le président américain hésite depuis des mois sur la stratégie globale à mener vis-à-vis des Européens durant l'année électorale 2020.

Trump révisé depuis deux années d'en découdre avec l'industrie automobile allemande. Il invoque régulièrement le déséquilibre entre les droits de douane américains (2,5 %), bien inférieurs aux 10 % infligés par les Européens sur les voitures outre-Atlantique. Quitte à oublier qu'une grande partie des voitures allemandes vendues aux Etats-Unis sont assemblées, voire fabriquées en Amérique du Nord.

Un armistice avait été signé entre Américains et Européens à l'été 2018, mais l'administration Trump veut sans cesse remettre à l'ordre du jour le dossier automobile et y ajouter celui de l'agriculture (privés de débouchés en Chine, les Français veulent écouler leur production en Europe, mais la France s'y oppose). M. Trump est censé prendre des décisions sur le sujet d'ici au 13 novembre, pour savoir s'il taxe unilatéralement 100 milliards de dollars d'exportations européennes de biens et de services. Celles-ci ont atteint, en 2018, 68,4 milliards de dollars, pour seulement 57,5 milliards d'importations en provenance des Etats-Unis.

JULIE BOUISSEAU, GUY DUTHIEL ET ARNAUD LEPARMENTIER

## LES DATES

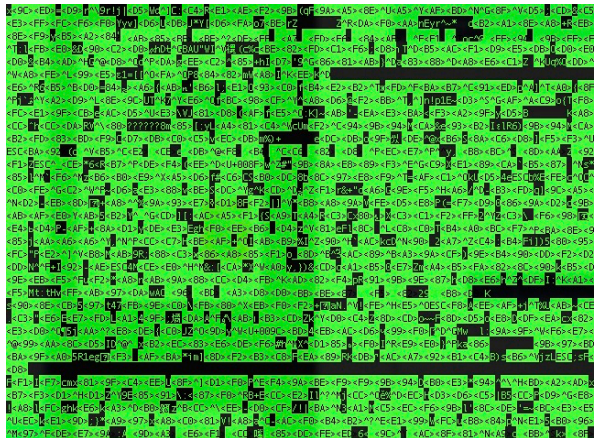
**2004**  
Boeing dépose plainte en octobre et demande à l'OMC d'enquêter sur les aides publiques versées par l'Allemagne, la France, le Royaume-Uni et l'Espagne à Airbus depuis son lancement en 1969.

**2005**  
Au mois de juin, Airbus se retourne à son tour contre Boeing, l'accusant de recevoir des subventions sous la forme d'abattements fiscaux et de déductions fiscales.

**2010**  
En juin, l'OMC conclut que les subventions versées à Airbus portent préjudice aux intérêts américains. L'avionneur européen fait appel. En septembre, un rapport préliminaire de l'OMC déclare que les subventions accordées à Boeing ont violé les règles de l'OMC et doivent être retirées.

**2018**  
L'organe d'appel de l'OMC confirme, en mai, que les subventions européennes versées à Airbus ont porté un préjudice commercial à Boeing et constate que l'avionneur européen ne s'est pas mis en conformité.

# STRUCTURE LOGIQUE ET CODAGE



Depuis 40 ans, nos commissions bâillonnent de nouveaux mondes

**finansol & argent**  
organisent

**LES GRANDS PRIX DE LA FINANCE SOLIDAIRE**

**2019**

**VOTEZ POUR LE PRIX DES TERRITOIRES**

Les nominés 2019 sont

**Ecoloc**, le lieu de vie qui redynamise son territoire rural.  
Valée de la Méouge

**Share-Wood**, l'atelier de menuiserie partagé qui réunit amateurs et professionnels.

**PERTES & PROFITS** | BOEING-AIRBUS  
PAR PHILIPPE ESCANDE

**Je te tiens, tu me tiens...**

Depuis qu'ils se font face, ces deux-là excellent dans l'interminable jeu de la barbotine. Le te tiens, tu me tiens... Sans autre rival sur l'aviation long-courrier, ils se cherchent sur les produits et les clients, mais aussi sur le tapis vert. Depuis quinze ans, Boeing reproche à son concurrent d'avoir bénéficié de généreuses subventions de la part de

d'avoir été soutenu par la puissance publique, grâce à de généreux contrats avec la défense nationale. La décision de l'OMC, qui valait sur la longue durée, ne pouvait donc tomber en 2020 et se compter également en milliards de dollars.

**Concurrence chinoise**  
L'effacement du dragon en l'air

Boeing. Si des droits de douane conséquents sont imposés de part et d'autre, une bonne partie de ces commandes tomberont à l'eau, pénalisant à la fois constructeurs et clients. De plus, les chaînes de valeur de Boeing et d'Airbus sont largement mondialisées, avec des sous-traitants dans tous les pays. Airbus achète pour 10 milliards de dollars de



# ALLER AU-DELÀ DES MOTS CLEFS : PAS TOUJOURS FACILE...

Google Scholar search results for "neural network". The search bar contains "neural network" and the results show approximately 265,000 articles. The left sidebar includes filters for date (indifferent, 2019, 2018, 2015, specific period), sorting (relevance, date), languages (French), and options to include patents/citations and create alerts.

Google Scholar search results for "neural networks". The search bar contains "neural networks" and the results show approximately 281,000 articles. The left sidebar includes filters for date (indifferent, 2019, 2018, 2015, specific period), sorting (relevance, date), languages (French), and options to include patents/citations and create alerts.

## Date indifférente

- Depuis 2019
- Depuis 2018
- Depuis 2015
- Période spécifique...

## Trier par pertinence

Trier par date

## Toutes les langues

Rechercher les pages en Français

- inclure les brevets
- inclure les citations

Créer l'alerte

## Neural network ensembles

[LK Hansen, P Salamon](#) - IEEE Transactions on Pattern Analysis & ..., 1990 - computer.org  
ABSTRACT<p> Several means for improving the performance and training of **neural networks** for classification are proposed. Crossvalidation is used as a tool for optimizing **neural network** parameters and architecture. It is shown that the remaining residual generalization ...  
☆ 📄 Cité 4055 fois [Autres articles](#) [Les 20 versions](#)

## Neural network-based face detection

[HA Rowley, S Baluja, T Kanade](#) - IEEE Transactions on pattern ..., 1998 - ieeexplore.ieee.org  
We present a **neural network**-based upright frontal face detection system. A retinally connected **neural network** examines small windows of an image and decides whether each window contains a face. The system arbitrates between multiple networks to improve ...  
☆ 📄 Cité 5399 fois [Autres articles](#) [Les 67 versions](#)

## [PDF] A general regression neural network

[DF Specht](#) - IEEE transactions on **neural networks**, 1991 - Citeseer  
This paper describes a memory-based **neural network** that provides estimates of continuous variables and converges to the underlying (linear or nonlinear) regression surface. This general regression **neural network** (GRNN) is a one-pass learning algorithm with a highly ...  
☆ 📄 Cité 4039 fois [Autres articles](#) [Les 13 versions](#) 🔗

## Recurrent neural network based language model

[T Mikolov, M Karafiát, L Burget, J Černocký](#)... - ... annual conference of ..., 2010 - isca-speech.org  
A new recurrent **neural network** based language model (RNN LM) with applications to speech recognition is presented. Results indicate that it is possible to obtain around 50% reduction of perplexity by using mixture of several RNN LMs, compared to a state of the art ...  
☆ 📄 Cité 3364 fois [Autres articles](#) [Les 7 versions](#)

## Recherches associées

**artificial** neural network      neural network **prediction**  
**convolutional** neural network      **fuzzy** neural network  
**recurrent** neural network      neural network **matlab**  
**deep** neural network      neural network **forecasting**

## Date indifférente

- Depuis 2019
- Depuis 2018
- Depuis 2015
- Période spécifique...

## Trier par pertinence

Trier par date

## Toutes les langues

Rechercher les pages en Français

- inclure les brevets
- inclure les citations

Créer l'alerte

## [LIVRE] Neural networks: a comprehensive foundation

[S Haykin](#) - 1994 - dl.acm.org  
This book represents the most comprehensive treatment available of **neural networks** from an engineering perspective. Thorough, well-organized, and completely up to date, it examines all the important aspects of this emerging technology, including the learning ...  
☆ 📄 Cité 50331 fois [Autres articles](#) [Les 8 versions](#) 🔗

## [LIVRE] Neural networks and learning machines/Simon Haykin.

[SS Haykin](#) - 2009 - repository.fue.edu.eg  
Pearson Prentice Hall. All rights reserved. Printed in the United States of America. This publication is protected by Copyright and permission should be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any ...  
☆ 📄 Cité 6152 fois [Autres articles](#) [Les 2 versions](#) 🔗

## Imagenet classification with deep convolutional neural networks

[A Krizhevsky, I Sutskever, GE Hinton](#) - Advances in **neural** ..., 2012 - papers.nips.cc  
We trained a large, deep convolutional **neural network** to classify the 1.3 million high-resolution images in the LSVRC-2010 ImageNet training set into the 1000 different classes. On the test data, we achieved top-1 and top-5 error rates of 39.7% and 18.9% which is ...  
☆ 📄 Cité 47095 fois [Autres articles](#) [Les 95 versions](#) 🔗

## Reducing the dimensionality of data with neural networks

[GE Hinton, RR Salakhutdinov](#) - science, 2006 - science.sciencemag.org  
High-dimensional data can be converted to low-dimensional codes by training a multilayer **neural network** with a small central layer to reconstruct high-dimensional input vectors. Gradient descent can be used for fine-tuning the weights in such "autoencoder" **networks** ...  
☆ 📄 Cité 10174 fois [Autres articles](#) [Les 20 versions](#)

## Recherches associées

**artificial** neural networks      **fuzzy** neural networks  
**convolutional** neural networks      neural networks **delays**  
**deep** neural networks      neural networks **forecasting**  
**recurrent** neural networks      neural networks **time-varying delays**

Google Scholar

arbre feuille

Articles Environ 108000 résultats (0,06 s)

Date indifférente  
 Depuis 2019  
 Depuis 2018  
 Depuis 2015  
 Période spécifique...

Trier par pertinence  
 Trier par date

Toutes les langues  
 Rechercher les pages en Français

inclure les brevets  
 inclure les citations

Créer l'alerte

[LIVRE] **Arbres, arbustes et lianes des zones sèches d'Afrique de l'Ouest**  
 M Arbonnier - 2009 - books.google.com  
 ... à la richesse de leur vocabulaire puisque, pour une seule espèce, il peut y avoir un nom spécifique pour désigner l'**arbre**, sa **feuille**, sa fleur ... En revanche, en milieu fermé — galerie forestière, forêt claire —, où la concurrence avec les **arbres** voisins est grande, l'**arbre** a ten ...  
 ☆ 99 Cité 687 fois Autres articles Les 13 versions

[LIVRE] **Les arbres du Canada**  
 JL Farrar - 1996 - books.google.com  
 ... résinifère nervure médiane Chaque description de genre débute par quelques renseignements généraux : importance du genre, forme des **arbres** et région(s) ... **Feuilles** La **feuille** est l'organe que la plupart des lecteurs examineront tout d'abord pour identifier un **arbre** ...  
 ☆ 99 Cité 92 fois Autres articles Les 4 versions

[PDF] **Un procédé itératif de dénombrement d'arbres binaires et son application à leur génération aléatoire**  
 JL Rémy - RAIRO. Informatique théorique, 1985 - numdam.org  
 ... (A, v) = DÉFINITION 4 (**arbres** binaires pointés) : Un **arbre** binaire pointé A\* est la donnée d'un **arbre** binaire à **feuilles** numérotées A', d'un nœud ou d'une **feuille** x de cet **arbre** et d'une marque ce (g, d). Le nombre B\*n d'**arbres** binaires pointés de taille n est égal à  $2^{(2n-1)}B$  ...  
 ☆ 99 Cité 121 fois Autres articles Les 3 versions

[LIVRE] **Les arbres de connaissances**  
 M Authier, L Pierre - 2017 - books.google.com  
 ... On verra tout au long de l'ouvrage que le système des **arbres** de connaissances respecte les individus, n'impose rien, fonctionne sur ... La petite fille sort de son cartable une grande **feuille** qu'elle ... On peut y voir en couleurs un **arbre** aux multiples branches et au beau feuillage vert ...  
 ☆ 99 Cité 215 fois Autres articles

[PDF] **Arbres de décision**  
 R Rakotomalala - Revue Modulad, 2005 - modulad.fr  
 ... d'utiliser des règles plus efficaces pour définir la taille adéquate de l'**arbre** de décision ... lorsqu'elle est pure, quelle est la règle de décision optimale lorsque qu'une **feuille** contient des ... Répondre à ces questions permet de définir une méthode d'induction des **arbres** de décision ...  
 ☆ 99 Cité 79 fois Autres articles Les 10 versions

**Les glycosides flavoniques foliaires de quelques arbres, au cours du cycle végétatif**  
 M Tissut, K Egger - Phytochemistry, 1972 - Elsevier  
 ... Les glycosides flavoniques foliaires de quelques **arbres** ... complexes qui libèrent du glucosyl-3 kaempferol par hydrolyse alcaline; les flavonols de la **feuille** adulte sont ... glucosides et les 3-arabinosides apparaissent sur les chromatogrammes des extraits de **feuilles** jeunes, les 3 ...  
 ☆ 99 Cité 30 fois Autres articles Les 2 versions

# CONTEXTES ET CONNAISSANCES

Indexation par mots, par terme, par domaine...

Google Scholar

méthane

Articles Environ 61100 résultats (0,08 s)

Date indifférente  
 Depuis 2019  
 Depuis 2018  
 Depuis 2015  
 Période spécifique...

Trier par pertinence  
 Trier par date

Toutes les langues  
 Rechercher les pages en Français

inclure les brevets  
 inclure les citations

Créer l'alerte

**Methane emissions from cattle**  
 KA Johnson, DE Johnson - Journal of animal ecology  
 Increasing atmospheric concentrations of methane from ruminants are a major source of origin. Ruminant livestock can produce methane from the rumen. The level of production results in estimates of the methane emissions from cattle.  
 ☆ 99 Cité 2309 fois Autres articles

**Biogeochemical aspects of atmospheric methane**  
 RJ Cicerone, RS Oremland - Global biogeochemical cycles and air quality  
 Methane is the most abundant organic chemical in the atmosphere. Its concentration is increasing with time, as a variety of independent processes contribute to its increase. Photochemical reactions oxidize methane in the atmosphere.  
 ☆ 99 Cité 1901 fois Autres articles

**Methane production by ruminants**  
 AR Moss, JP Jouany, J Newbold - Annales de la recherche vétérinaire  
 The aim of this paper is to review the role of methane production by ruminants in the climate. The contribution of methane to the greenhouse effect is discussed. Agricultural emissions of methane from ruminants are reviewed.  
 ☆ 99 Cité 1152 fois Autres articles

CH4

Environ 1160000 résultats (0,08 s)

**PRODUCTION OF CO AND CH4, CO2 AT METAL ELECTRODES IN AQUEOUS SOLUTION**  
 Y Hori, K Kikuchi, S Suzuki - Chemistry Letters  
 The total analysis of gaseous and soluble products from the reduction of CO2 (current density 5 mA cm<sup>-2</sup>) was predominantly produced at Cd, In, Sn, and Pb electrodes.  
 ☆ 99 Cité 707 fois Autres articles

**Soil microorganisms as controllers of atmospheric methane (CH4, CO2, and NO).**  
 R Conrad - Microbiol. Mol. Biol. Rev., 1996 - microbiology  
 Production and consumption processes in soil and the atmosphere of methane (CH4), carbon dioxide (CO2), and nitrous oxide (NO) are discussed. Soil microbial processes contribute to the greenhouse effect.  
 ☆ 99 Cité 1758 fois Autres articles

**CO2 Reforming of CH4**  
 MCJ Bradford, MA Vannice - Catalysis Reviews: Science and Technology  
 Although technological practice should minimize methane emissions, the use of methane as a feedstock for the production of syngas is increasing.  
 ☆ 99 Cité 1152 fois Autres articles

# 4

## MULTILINGUISME, CITATIONS, FORMULES, RESULTATS...

OpenEdition : OpenEdition Books OpenEdition Journals Calenda Hypothèses Lettre OpenEdition Freemium

DOI / Références Télécharger

Accueil > Numéros > 21 > DOSSIER THÉMATIQUE Strangers at ... > Ἰν'ἀμνήμων τύχη γένοιτο πολλῶν δε...

# γαῖα

Revue interdisciplinaire sur la Grèce archaïque

Recherche

21 | 2018  
Varia

DOSSIER THÉMATIQUE  
Strangers at Home. Civilizing Immigrants between Inclusion and Exclusion in Ancient Thebes

**Ἰν'ἀμνήμων τύχη γένοιτο πολλῶν δεομένη σοφισμάτων (Phéniciennes, 64-65). La souveraineté brisée de la famille d'Œdipe et la crise de la parole dans le mythe tragique des Phéniciennes**

Ἰν'ἀμνήμων τύχη γένοιτο πολλῶν δεομένη σοφισμάτων (Phoenician Women, 64–65).  
*The Broken Sovereignty of Oedipus' Family and the Crisis of Speech in the Tragic Myth of Phoenician Women*

Ἰν'ἀμνήμων τύχη γένοιτο πολλῶν δεομένη σοφισμάτων (Fenicie, 64-65). *La sovranità spezzata della famiglia di Edipo e la crisi della parola nel mito tragico delle Fenicie*

AVEZZÙ Guido, *Il mito sulla scena. La tragedia ad Atene*, Venise, Marsilio, 2003.

AVEZZÙ Guido, « Emulazione e antagonismo nella produzione tragica ateniese », *Dionysus ex machina*, 6 (1), 2015, p. 137-156.

BATTEZZATO Luigi, « An Introduction to Tragedy. G. Avezù: *Il mito sulla scena. La tragedia ad Atene* », *The Classical Review*, 55 (1), 2005, p. 29.

DOI : [10.1093/clrevj/bni019](https://doi.org/10.1093/clrevj/bni019)

BEARZOT Cinzia, « Perdonare il traditore? La tematica amnistiale nel dibattito sul richiamo di Alcibiade », dans M. Sordi (éd.), *Amnistia, perdono e vendetta nel mondo antico*, Milan, Vita e Pensiero, 1997, p. 29-52.

BELTRAMETTI Anna, « Antigone o della questione morale. Elaborazione tragica della sovranità democratica », dans D. Ambaglio (éd.), « Συγγραφή ». *Materiali e appunti per lo studio della storia e della letteratura antica*, vol. 4, Como, Edizioni New Press, 2002, p. 33-49.



Table 6. The NDCG and Hit Ratio (HT) Results of AARM and Its Variants on Five Datasets for RQ3

Measures (%)	Movies		CDs		Clothings		Cell Phones		Beauty	
	NDCG	HT	NDCG	HT	NDCG	HT	NDCG	HT	NDCG	HT
AARM	<b>5.020</b>	<b>15.187</b>	7.252	<b>20.749</b>	1.957	<b>4.915</b>	<b>4.976</b>	<b>11.568</b>	<b>5.314</b>	<b>13.648</b>
A_Static	4.376	13.318	6.794	19.567	1.898	4.590	4.728	11.181	4.918	12.735
No-UserAtt	4.290	13.104	6.700	19.108	1.310	3.217	4.685	10.786	4.739	12.297
Impr A_static	14.717	14.034	6.741	6.041	3.109	7.081	5.245	3.461	8.052	7.169
Impr No-UserAtt	17.016	15.896	8.239	8.588	49.389	52.782	6.211	7.250	12.133	10.986

We follow the short form convention adopted in Table 4 to name the datasets. The best performance of each measure on each dataset is highlighted in bold. The last block shows the percentage of improvements (or decrements for negative values) achieved by AARM compared with A\_static (Impr A\_static) and No-UserAtt (Impr No-UserAtt).

Table 7. The Corresponding Precision and Recall Results of AARM and Its Variants on Five Datasets for RQ3

## 5 CONCLUSION AND FUTURE WORK

In this article, we presented an AARM, which carefully captures the interactions between extracted from reviews for recommendation. AARM first calculates the interactions between aspect embeddings to estimate how a product fits a user's requirements on each aspect, and then estimates the user's overall satisfaction on the product by synthesizing the product's performance on each aspect. To deal with the problem that the number of shared aspects between a user and a product is often limited, AARM takes the interactions between different aspects into consideration. With a well-designed aspect-level attention module, not only the shared aspects but also other related aspect pairs can be selected and assigned higher attention values. In addition, we hold the assumption that a user's interests toward aspects are varied when examining different products. To achieve the goal, an attention module which simultaneously considers user and product information is designed in AARM. In the experiments on five real-world datasets, AARM outperforms the state-of-the-art methods on the top-N recommendation task. In particular, compared with multimodal (textual reviews, product images, and numerical ratings) methods JRL and eJRL, AARM can still achieve better results in all datasets. To demonstrate the effectiveness of each component in AARM, a lot of quantitative experiments and qualitative case studies are conducted.

Figure 3 shows an exemplary 6-step route for an intermediate of a drug candidate synthesis reported in 2015, which has been found by our algorithm in 5.4 s. It matches the published route.[45]

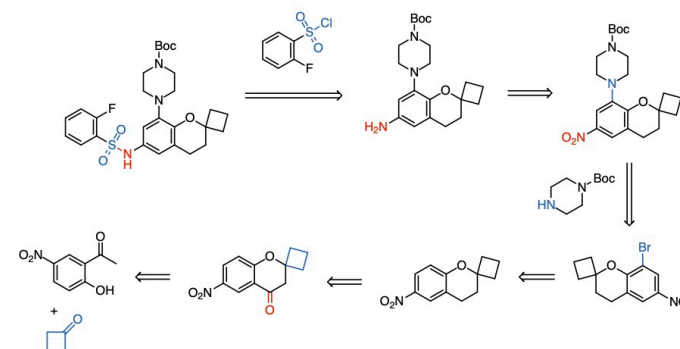


Figure 3: An exemplary 6-step route for an intermediate of a drug candidate synthesis reported in 2015, which has been found by our algorithm in 5.4 s. It matches the published route.[45]

Figure 3: An exemplary 6-step route for an intermediate of a drug candidate synthesis reported in 2015, which has been found by our algorithm in 5.4 s. It matches the published route.[45]

$$\mathbf{x}_{v,i} = \mathbf{g}_v \odot \mathbf{c}_i,$$

$$\mathbf{g}_v = \sum_{j \in A_v} \mathbf{c}_j.$$

interaction between two aspects represents their similarity,  $\mathbf{x}_{v,i}$  represents the interaction between the aspect  $a_i$  and the product  $v$ . To measure the importance of aspect  $a_i$ ,  $\mathbf{x}_{v,i}$  is used as aspect  $a_i$ 's input to the user-level attention layer as

$$\hat{\alpha}_{u,v,i} = \mathbf{w}_{att_2}^T \mathbf{x}_{v,i},$$

$$\alpha_{u,v,i} = \frac{\exp(\hat{\alpha}_{u,v,i})}{\sum_{j \in A_u} \exp(\hat{\alpha}_{u,v,j})}.$$

$\mathbf{w}_{att_2} \in \mathbb{R}^{d_a}$  is a learnable vector, and  $\alpha_{u,v,i}$  represents the importance of aspect  $a_i$  with regard to product  $v$ . This attention layer is different





## 2 L'EXISTANT

De nombreux acteurs

1

## DE NOMBREUX ACTEURS INDUSTRIELS

(et de très nombreux laboratoires publics)

xerox.fr  
Transformation digitale des entreprises - Xerox

À propos | Services | Produits | Fournitures | Support Client | Où acheter

Partager | Tweet | Lien | Courrier électronique | Imprimer | Plus

### Analyse documentaire : l'arme secrète pour être en première ligne de la transformation numérique

Pour la plupart des entreprises, les ambitions numériques restent lettres mortes, comme le montre une enquête Xerox réalisée auprès de responsables IT. Le désir d'abandonner le papier pour

Elsevier R&D Solutions  
FOR PHARMA & LIFE SCIENCES

## Pathway Studio®

Fact Sheet

Linguamatics  
an IQVIA company

Technology | Products | Solutions



### AI Siblings: NLP and Machine Learning for Better Drug Discovery

SINEQUA

Insight Platform | Industry Focus | Use Cases | Resources | Customers | Partners | Company | News | Blog

## Instant Insights in Your Work Content

With Cognitive Search and Powerful Analytics

Synapse

BESOINS | SOLUTIONS | PROJETS

## SYNAPSE DÉVELOPPEMENT

Les experts de l'Intelligence Artificielle appliquée au texte

MONDECA

MAKING SENSE OF CONTENT

## COMPRENDRE ET INTERPRETER

Comprendre le texte, l'image, la vidéo. Identifier ce qui est significatif. Catégoriser. Évaluer. Trouver des relations. Désambiguïser. Enrichir. Annoter. Gérer la connaissance pour nourrir les solutions d'intelligence artificielle: Classification automatique, bots, gestion d'alertes.

syllabs

Accueil | Immobilier | Nos offres d'emploi | Contact | EN | ES

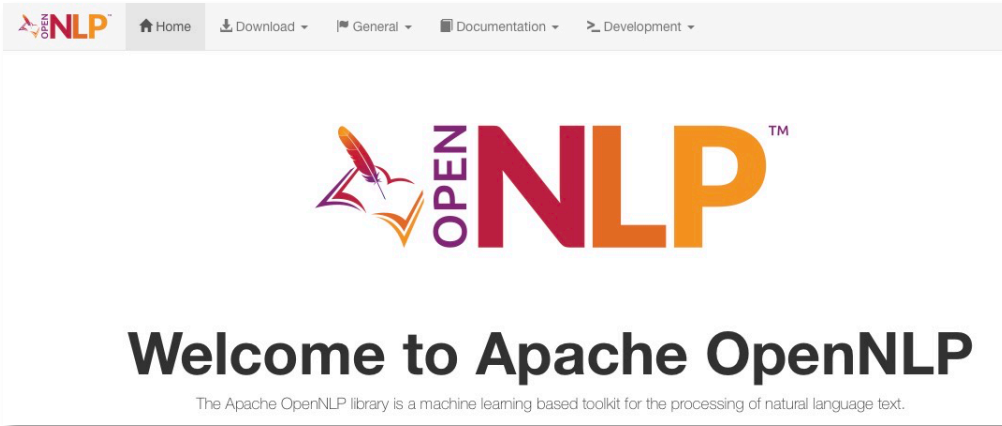
## Entrez dans le futur de la création de contenu.

Syllabs propose des solutions automatisées de création de textes et d'optimisation de



# 2

## DES PROJETS OUVERTS



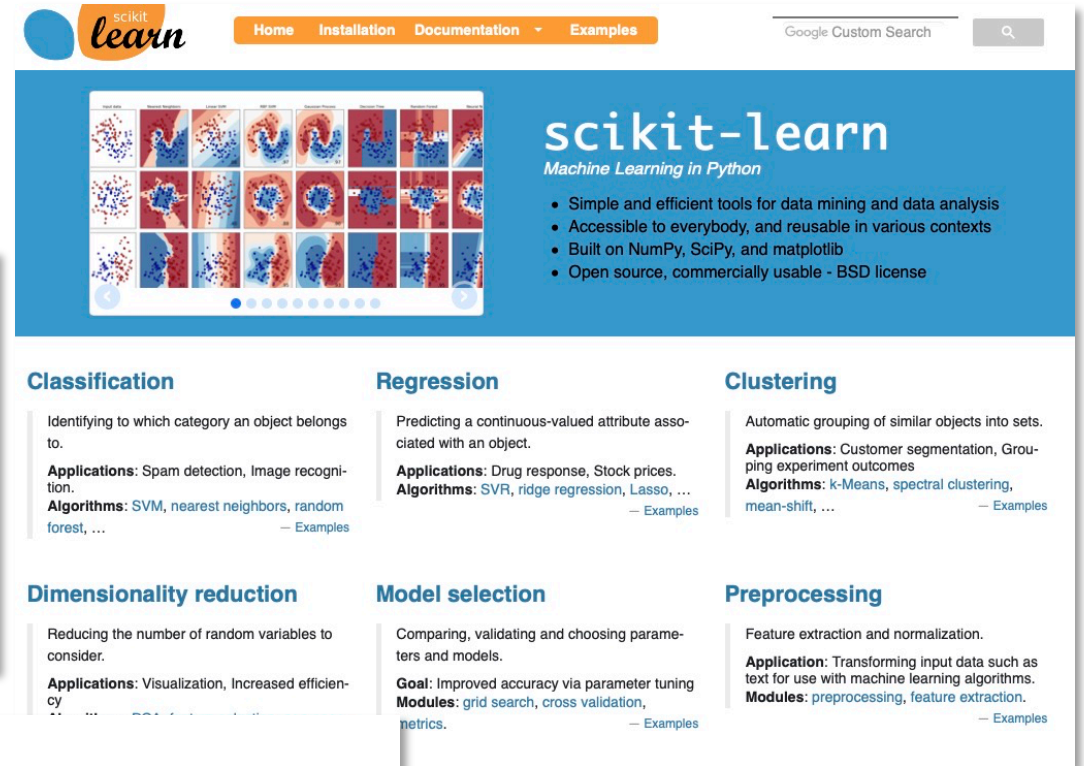
OPEN NLP

Home Download General Documentation Development

# OPEN NLP™

## Welcome to Apache OpenNLP

The Apache OpenNLP library is a machine learning based toolkit for the processing of natural language text.



scikit-learn

Home Installation Documentation Examples

Google Custom Search

# scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

### Classification

Identifying to which category an object belongs to.  
**Applications:** Spam detection, Image recognition.  
**Algorithms:** SVM, nearest neighbors, random forest, ...

### Regression

Predicting a continuous-valued attribute associated with an object.  
**Applications:** Drug response, Stock prices.  
**Algorithms:** SVR, ridge regression, Lasso, ...

### Clustering

Automatic grouping of similar objects into sets.  
**Applications:** Customer segmentation, Grouping experiment outcomes  
**Algorithms:** k-Means, spectral clustering, mean-shift, ...

### Dimensionality reduction

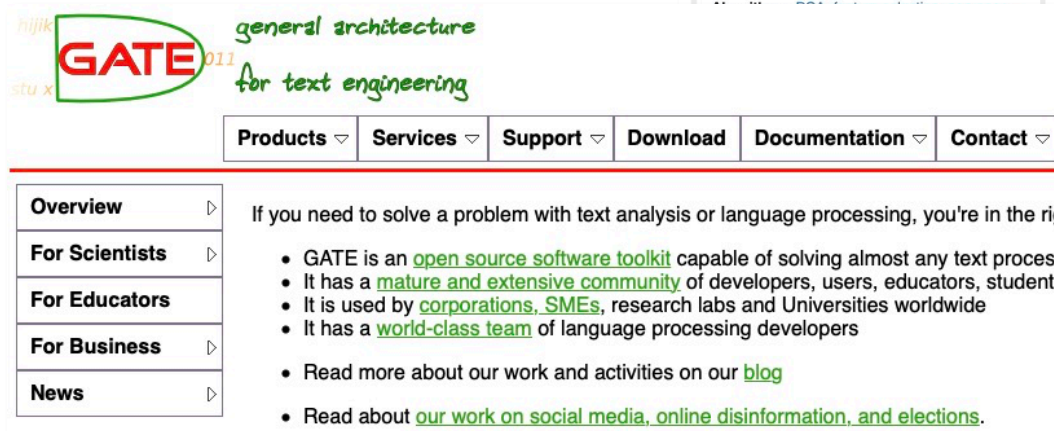
Reducing the number of random variables to consider.  
**Applications:** Visualization, Increased efficiency

### Model selection

Comparing, validating and choosing parameters and models.  
**Goal:** Improved accuracy via parameter tuning  
**Modules:** grid search, cross validation, metrics.

### Preprocessing

Feature extraction and normalization.  
**Application:** Transforming input data such as text for use with machine learning algorithms.  
**Modules:** preprocessing, feature extraction.



hijkl stu x

# GATE 011

general architecture for text engineering

Products Services Support Download Documentation Contact

### Overview

If you need to solve a problem with text analysis or language processing, you're in the right place.

### For Scientists

- GATE is an **open source software toolkit** capable of solving almost any text processing problem
- It has a **mature and extensive community** of developers, users, educators, students
- It is used by **corporations, SMEs**, research labs and Universities worldwide
- It has a **world-class team** of language processing developers

### For Educators

- Read more about our work and activities on our **blog**

### For Business

- Read about **our work on social media, online disinformation, and elections.**

### News

# 2

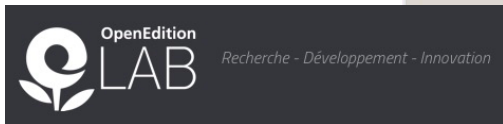
## DES LOGICIELS OUVERTS POUR L'IST : EXEMPLE *BILBO*

The screenshot shows the website 'cybergegeo' (European Journal of Geography). The main article is titled '643 Modélographie multi-agents de la simulation des interactions sociétés-environnement et de l'émergence des villes' by Clara Schmitt et Denise Pumain. A red box highlights the 'Bibliographie' link in the article's navigation menu. A red arrow points from this link to a list of references in a separate window. The references include:

- Banos A., Sanders L., 2012, « Vers une cartographie sémantique des modèles urbains : des individus aux systèmes de villes », *colloque Modélisation urbaine : de la représentation au projet*, Coll. Références, Commissariat Général au Développement Durable.
- Bretagnolle A., Pumain D., 2010, « Simulating urban networks through multiscale space-time dynamics (Europe and United States, 17th -20th centuries) », *Urban Studies*, vol. 47, No.13, 2819-2839. DOI : 10.1177/0042098010377366
- Bura S., Guerin-Pace F., Mathian H., Pumain D., Sanders L., 1996, « Multiagent systems and the dynamics of a settlement system », *Geographical Analysis*, vol. 28, No.2, 161-178. DOI : 10.1111/j.1538-4632.1996.tb00927.x
- Crooks A., Castle C., Batty M., 2008, « Key challenges in agent-based modelling for geo-spatial simulation », *Computers, Environment and Urban Systems*, vol. 32, No.6, 417-430. DOI : 10.1016/j.compenvurbsys.2008.09.004

<https://lab.hypotheses.org/category/bilbo-bibliographical-robot>

Lier et citer les publications scientifiques.



# 2

## UNE INITIATIVE PUBLIQUE : ISTE<sup>X</sup>

Un moteur de recherche en texte intégral, une chaîne de traitements, des enrichissements, des APIs et des services



<https://www.istex.fr/>



23 millions de documents  
9 279 revues

autisme

Affiner votre recherche

Corpusname

- Edp-sciences 196
- Elsevier 129
- Wiley 109
- Springer-journals
- Cambridge 45 105
- Sage 23
- Lavoisier 16
- Bmj 8
- Springer-ebooks 7
- Brill-hacco 4

Categories.inist

- 1 - sciences humaines et so... 92
- 1 - sciences appliquées, te... 52
- 2 - sciences biologiques et... 52
- 3 - sciences médicales 52
- 2 - art et archéologie 5
- 2 - linguistique 5

Announcement

Pas de résumé disponible pour cet article

Journal of Autism and Developmental Disorders vol.23, page 427 - 427

Fulltext Metadata Enrichissements

**Harnessing Different Dimensions of Space: The Built Environmen**

Abstract: An understanding of diversity is a key principle in the development of theories, tools and techniques. Inclusive design, we want to gain a more accurate insight into the diversity o...  
S. Baumers. A. Heylighen.

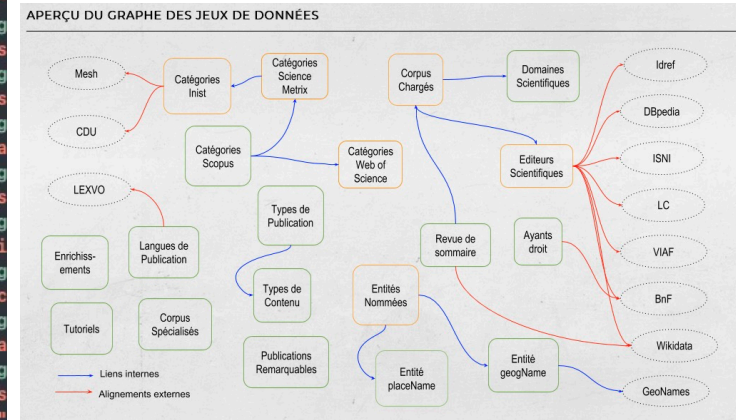
Designing Inclusive Interactions vol., page 13 - 23

Fulltext Metadata Enrichissements

**Sclérose tubéreuse de Bourneville et autisme**

Résumé: La sclérose tubéreuse de Bourneville comporte un grand polymorphisme dans sa symptomatologie et du comportement incluant des syndromes autistiques caractéristiques ou partiels. Les di...

```
.0-alpha" resp="istex-rd"
r"><term>orient</term><fs type="statistics"><f name="
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>Paris</term><fs type="statis
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>Nancy</term><fs type="statis
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>Gillberg</term><fs type="sta
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>Lagos</term><fs type="statis
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>France</term><fs type="stati
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>YS</term><fs type="statistic
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>New-York</term><fs type="sta
lock><annotationBlock corresp="text"><placeName chang
ity.data.istex.fr"><term>Tours</term><fs type="statis
lock>/listAnnotation<listAnnotation type="persName"
0-alpha" resp="istex-rd" scheme="https://persname-ent
ency"><numeric value="1"/></f></fs></persName></annot
```

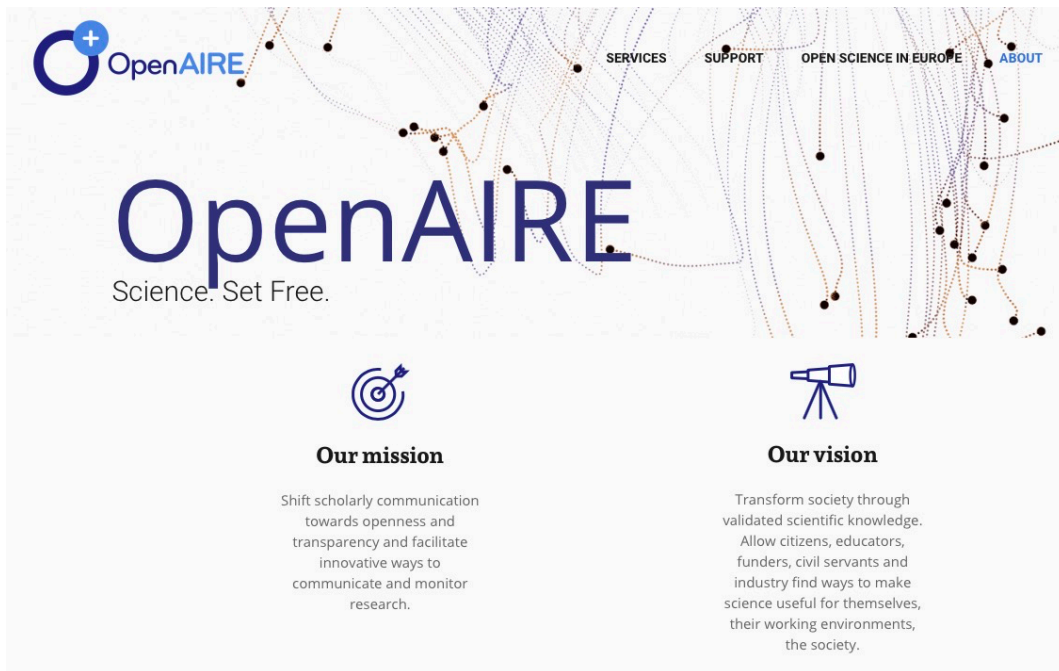


<https://data.istex.fr>



# 2

## DES INITIATIVES POUR UN COUPLE IST / TDM OUVERT



The image shows the OpenAIRE website header and two main sections: 'Our mission' and 'Our vision'. The header features the OpenAIRE logo, navigation links for SERVICES, SUPPORT, OPEN SCIENCE IN EUROPE, and ABOUT, and the tagline 'Science. Set Free.' Below this, 'Our mission' is represented by a target icon and text about shifting scholarly communication. 'Our vision' is represented by a telescope icon and text about transforming society through validated scientific knowledge.

**OpenAIRE**  
Science. Set Free.

SERVICES SUPPORT OPEN SCIENCE IN EUROPE ABOUT

**Our mission**  
Shift scholarly communication towards openness and transparency and facilitate innovative ways to communicate and monitor research.

**Our vision**  
Transform society through validated scientific knowledge. Allow citizens, educators, funders, civil servants and industry find ways to make science useful for themselves, their working environments, the society.

<https://www.openaire.eu>



The image shows the CORE website header and three main features. The header includes the CORE logo and the text 'The world's largest collection of open access research papers'. Below this is a search bar with the text 'Search 135,539,113 papers around the world' and a 'Search' button. Three icons represent the website's features: a globe for 'We aggregate and enrich open access research papers from around the world', a laptop for 'We provide seamless access to content and data, through our unique APIs', and a building for 'We create powerful services for researchers, universities, and industry'. A world map at the bottom shows the global reach of the collection.

**CORE**  
The world's largest collection of open access research papers

Search 135,539,113 papers around the world Search

We aggregate and enrich open access research papers from around the world  
[Read about our data](#)

We provide seamless access to content and data, through our unique APIs  
**Perfect for text mining!**

We create powerful services for researchers, universities, and industry



## Project mining details

Provide your **UTF-8** encoded text, on the current URL using the HTTP POST method.

You may also choose among the available mining processes.

HTTP POST parameters:

- **document:** UTF-8 encoded text
- **projects:** Project processing (on/off)
- **data citations:** Data citation processing (on/off)
- **classification:** Classification processing (on/off)

The service will return a **JSON** encoded result containing the following fields:

- **funding\_info:** Result category
- **fund:** Funder name (e.g., *FP7*, *Wellcome Trust*)
- **acronym:** The project acronym (only for FP7 projects)
- **grantid:** The project grant identifier
- **confidence:** Confidence weight

Result example:

```
{"funding_info": [{"fund": "fp7", "acronym": "CORONET", "grantid": "269459", "confidence": 0.96}]}
```

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Ut enim ad minim veniam, quis nostrud exercitation ullamco laboris nisi ut aliquip ex ea commodo consequat. Duis aute irure dolor in reprehenderit in voluptate velit esse cillum dolore eu fugiat nulla pariatur. Excepteur sint funded by fp7 project OpenAIREplus occaecat cupidatat non proident,fp7 project with grant agreement number 318338 sunt in culpa qui officia deserunt mollit anim id est laborum.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. In posuere felis nec tortor. Pellentesque faucibus. Ut accumsan ultricies elit. Maecenas at justo id velit placerat molestie. Donec dictum lectus non odio. Cras a ante vitae enim iaculis aliquam. Mauris nunc quam, venenatis nec, euismod sit amet, egestas placerat, est. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Cras id elit. Integer quis urna. Ut ante enim, dapibus malesuada, fringilla eu, condimentum quis, tellus. Aenean porttitor eros vel dolor. Donec convallis pede venenatis nibh. Duis quam. Nam eget lacus. Aliquam erat volutpat. Quisque dignissim congue leo.

Mauris vel lacus vitae felis vestibulum volutpat. Etiam est nunc, venenatis in, tristique eu, imperdiet ac, nisl. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. In iaculis facilisis massa. Etiam eu urna. Sed porta. Suspendisse quam leo, molestie sed, luctus quis, feugiat in, pede. Fusce tellus. Sed metus augue, convallis et, vehicula ut, pulvinar eu, ante. Integer orci tellus, tristique vitae, consequat nec, porta vel, lectus. Nulla sit amet diam. Duis non nunc. Nulla rhoncus dictum metus. Curabitur tristique mi condimentum orci. Phasellus pellentesque aliquam enim. Proin dui lectus, cursus eu, mattis laoreet, viverra sit amet, quam. Curabitur vel dolor ultrices ipsum dictum tristique. Praesent vitae lacus. Ut velit enim, vestibulum non, fermentum nec, hendrerit quis, leo. Pellentesque rutrum malesuada neque.

References:

Sieger, Rainer; (2012): PanGet - downloads multiple data sets from PANGAEA; PANGAEA - Data Publisher for Earth & Environmental Science.

Grobe, Hannes; (2005): Description and user manual of the information system PANGAEA; PANGAEA - Data Publisher for Earth & Environmental Science. <http://dx.doi.org/10.1594/PANGAEA.319947>

Appendix:

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. In posuere felis nec tortor. Pellentesque faucibus. Ut accumsan ultricies elit. Maecenas at justo id velit placerat molestie. Donec dictum lectus non odio. Cras a ante vitae enim iaculis aliquam. Mauris nunc quam, venenatis nec, euismod sit amet, egestas placerat, est. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Cras id elit. Integer quis urna. Ut ante enim, dapibus malesuada, fringilla eu, condimentum quis, tellus. Aenean porttitor eros vel dolor. Donec convallis pede venenatis nibh. Duis quam. Nam eget lacus. Aliquam erat volutpat. Quisque dignissim congue leo.

Projects  Data citations  Classification

```
{"funding_info": [{"fund": "EC", "acronym": "OPTIQUE", "grantid": "318338", "confidence": 0.92}], "classification_info": [], "datacitation_info": [{"organization": "datacite", "related_doi": "10.1594/PANGAEA.319947", "confidence": 1, "resource_type": "Dataset"}, {"organization": "datacite", "related_doi": "10.1594/PANGAEA.804562", "confidence": 0.69, "resource_type": "Dataset"}]}
```

# 2

## DES INITIATIVES EUROPEENNES : L'ERIC CLARIN



Members	National Consortium (NC)	Leading NC partner
Austria	Digital Humanities Austria	ACDH-OEAW
Bulgaria	CLADA-BG	Bulgarian Academy of Sciences
Croatia	HR-CLARIN	University of Zagreb
Cyprus	CLARIN-CY	Digital Heritage Research Lab (Cyprus University of Technology)
Czech Republic	LINDAT/CLARIN	Charles University Prague
Denmark	CLARIN-DK	University of Copenhagen
Estonia	CLARIN Estonia	Center of Estonian Language Resources
Finland	FIN-CLARIN	University of Helsinki
Germany	CLARIN-D	University of Tuebingen
Greece	clarin-el	ILSP-ATHENA Research Center
Hungary	HunCLARIN	Research Institute for Linguistics, Hungarian Academy of Sciences
Italy	CLARIN-IT	Institute for Computational Linguistics A. Zampolli, Italian National Research Council
Latvia	CLARIN-LV	Institute of Mathematics and Computer Science, University of Latvia
Lithuania	CLARIN-LT	Vytautas Magnus University
The Netherlands	CLARIAH-NL	Utrecht University
Norway	CLARINO	University of Bergen
Poland	CLARIN PL	Wroclaw University of Technology
Portugal	PORTULAN CLARIN	University of Lisbon
Slovenia	CLARIN-SI	Jožef Stefan Institute
Sweden	SWE-CLARIN	Språkbanken
Observer	National Consortium (NC)	Leading partner NC
France	Huma-Num	the National Center for Scientific Research (CNRS)
Iceland	CLARIN Iceland	The Árni Magnússon Institute for Icelandic Studies
South Africa	SADiLaR	North-West University
United Kingdom	CLARIN-UK	Oxford University



<https://www.clarin.eu>



# 2

## DES INITIATIVES EUROPÉENNES : L'INFRASTRUCTURE *OPENMINTED* 2015-2018

openMINTeD  
Open Mining Infrastructure for Text & Data

SERVICES JOIN ABOUT NEWS CONTACT US

**OpenMinTeD**  
**FOR RESEARCHERS**  
[services.openminded.eu](http://services.openminded.eu)



## Resource Type

Applications (41) 

## Refine

### Licence

- Affero General Public License v1.0 (1)
- BSD-2-Clause (Simplified) License (1)
- Creative Commons Attribution Share Alike 3.0 Unported (1)
- GNU Lesser General Public License v3.0 (1)
- Non standard Licence or Terms of use (2)
- [View more...](#)

### Rights Statement

- Open Access (41)

### Language

- Bulgarian (1)
- Catalan; Valencian (1)
- Czech (1)
- Danish (1)
- German (1)
- Spanish; Castilian (1)
- Irish (1)
- Galician (1)
- Croatian (1)
- Latvian (1)
- [View more...](#)

### Function

- Extraction of funding information

Showing 1 - 10 of 41 results

< Previous

PAGE 1 OF 5



## GeoPolitical Extractor App

Extracts geopolitical terms.



## IXA pipes for Basque for PDF files

IXA pipes for Basque with PDF reader. eu-ixa-pipes-omtd prov tokenizer, POS tagger, lemmatizer, NER tagger, Chunking and EC Classification. It reads from an input folder containing XML doc and outputs the added annotations in XML format to an output directory.



## Leica Model Annotation App

The application annotates a corpus with Leica Microsystems p



## MADIS FUNDING MINING

The Funding Mining application mines the fulltext of publicatip extracts links to projects. Currently, projects from EC (FP7/H2 (National Science Foundation, USA), NIH (National Institute of Health, USA), Wellcome Trust, FCT (Fundação para a Ciência e a Tecnologia Portugal), ARC (Australian Research Council), NHMRC (National Health and Medical Research Council, Australia), CSF/HRZZ (Hrvatska Znanost, Croatia), MSES-MZOS (Ministarstvo Znanosti, Obruke i Športa, Croatia), SFI (Science Foundation Ireland), NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek, Netherlands) are supported, but new funders are added regularly.

Recognizes phenotypes, genes, markers, and wheat-related taxa. It categorizes the phenotypes with the Wheat Trait Ontology.



## Alvis Arabidopsis Gene Regulation Extractor

Recognizes Gene, Protein and RNA of Arabidopsis thaliana. It normalizes them with Gene Locus and identifies interactions with relationships between Gene and Protein.



## Lancaster Stemmer (DKPro Core)

This Paice/Husk Lancaster stemmer implementation only works with the English language so far.



## Variable Disambiguator

Assign variable IDs to sentences based on calculating the similarity between the sentence text and the description of the variable.



## ClearNLP Segmenter (DKPro Core)

Tokenizer using Clear NLP.



## de-ixa-pipes-omtd

IXA pipes for German. It provides tokenizer, POS tagger, lemmatizer and NER tagger. It reads from an input folder containing XML documents and outputs the added annotations in XML format to an output directory.

### Function

- Analyzer (1)
- Annotator of semantic role labels (1)
- Chunker (1)
- Co-reference annotator (1)
- Constituency parser (1)
- Information extraction (1)
- Parser (1)
- Variables detector (1)
- Document classifier (2)
- [View more...](#)

### Component Distribution

- Medium**
- Executable Code (28)
- Docker Image (35)

### Processing Resource Type

- Lexical Conceptual Resource (6)
- Document (28)
- Corpus (29)

### Data Format

- Binary CAS (1)
- Binary format (1)
- CoNLL-U (1)
- CSV (1)
- GATE format (1)
- TSV (1)
- UIMA CAS format (1)
- ALVIS Enriched Document format (3)
- XML (4)
- [View more...](#)

### Annotation Type

- Citation (1)
- Constituent (1)
- Coreference (1)
- Dependency tree (1)
- Discourse annotation type (1)
- Event (1)
- Funding (1)
- Grape variety (1)

## CONCLUSION : LE TDM...

### Nécessite :

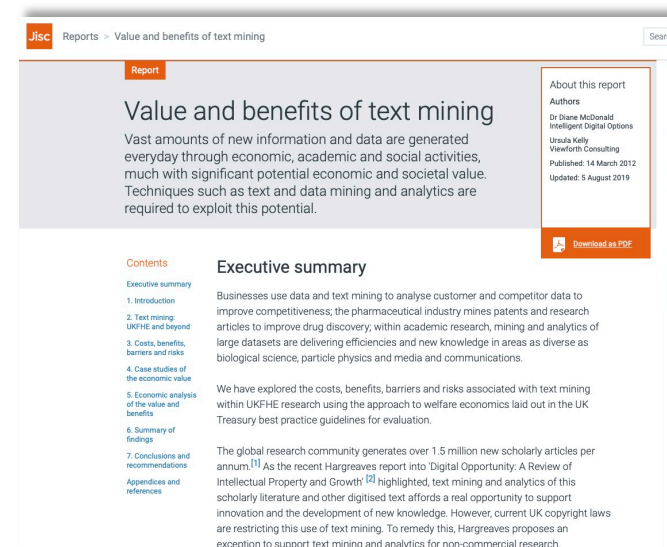
- un corpus cible, des ressources de spécialité
- d'intégrer différents composants logiciels
- un scénario et une référence pour évaluer la chaîne de traitements

### Faisable si :

- les composants sont interopérables, les métadonnées compatibles
- une infrastructure ouverte de services de fouille de textes est disponible...
- l'exception TDM sur la directive du droit d'auteur / copyright est soutenue

### Concerne et impacte :

- La recherche scientifique dans son ensemble
- La société au travers d'applications du quotidien



The screenshot shows the Jisc report page for 'Value and benefits of text mining'. The page includes a search bar, a 'Report' tab, and a 'Download as PDF' button. The main content area is divided into 'Contents' and 'Executive summary'. The 'Contents' section lists: 1. Introduction, 2. Text mining: UKFHE and beyond, 3. Costs, benefits, barriers and risks, 4. Case studies of the economic value, 5. Economic analysis of the value and benefits, 6. Summary of findings, 7. Conclusions and recommendations, and Appendices and references. The 'Executive summary' section states: 'Businesses use data and text mining to analyse customer and competitor data to improve competitiveness; the pharmaceutical industry mines patents and research articles to improve drug discovery, within academic research, mining and analytics of large datasets are delivering efficiencies and new knowledge in areas as diverse as biological science, particle physics and media and communications. We have explored the costs, benefits, barriers and risks associated with text mining within UKFHE research using the approach to welfare economics laid out in the UK Treasury best practice guidelines for evaluation. The global research community generates over 1.5 million new scholarly articles per annum [1]. As the recent Hargreaves report into 'Digital Opportunity: A Review of Intellectual Property and Growth' [2] highlighted, text mining and analytics of this scholarly literature and other digitised text affords a real opportunity to support innovation and the development of new knowledge. However, current UK copyright laws are restricting this use of text mining. To remedy this, Hargreaves proposes an exception to support text mining and analytics for non-commercial research.'

<https://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining>





Depuis 80 ans, nos connaissances  
bâtissent de nouveaux mondes

## MERCI DE VOTRE ATTENTION

# Visa TM Day le 15 novembre : vers une infrastructure de services avancés en text-mining

Dans le cadre du projet Visa TM du Comité pour la Science Ouverte, un « **Visa TM Day** » sera organisé **vendredi 15 novembre 2019** au ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation à **Paris**.

Le projet Visa TM a pour objectif l'étude d'une e-infrastructure de recherche pour la création d'une offre de service en fouille de textes pour la recherche, basée sur l'analyse sémantique et s'appuyant sur le potentiel de combinaison et d'adaptation offert par la plateforme européenne OpenMinTeD.

Autour de conférences et d'ateliers prospectifs, cette journée est destinée à dresser un état des lieux et discuter des perspectives concrètes ouvertes par les résultats du projet.

**Les inscriptions sont ouvertes jusqu'au 15 octobre.**

Programme et inscriptions sur <https://journées.inra.fr/visa-tm-day/>