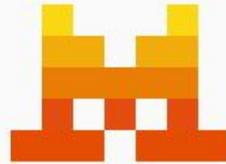


Usage de l'IA dans ISTEX

Journées Couperin
19 mars 2025



Demander au Chat



Demander au Chat

Qu'est-ce que l'IA à l'Inist-CNRS ?



Qu'est-ce que l'IA à l'Inist-CNRS ?



L'Inist-CNRS utilise des programmes informatiques pour **traiter, enrichir** des **documents numériques** pour faciliter la recherche, la lecture ou l'analyse.

IA sur des documents \approx fouille de textes, ou "Text and Data Mining - TDM".

Demander au Chat

Est-ce nouveau à l'Inist-CNRS ?



Est-ce nouveau à l'Inist-CNRS



En 1992, des personnes de l'Inist travaillaient déjà sur l'extraction de **"topics"**.

Hypertext paradigm in the field of information retrieval : a neural approach.

LELU Alain
ENST / Economie et Management
46 rue Barrault 75013 PARIS, FRANCE

FRANCOIS Claire
INSTITUT DE L'INFORMATION SCIENTIFIQUE ET TECHNIQUE (INIST/CNRS)
Département Recherche et Produits Nouveaux
2 allée du Parc de Brabois 54514 VANDOEUVRE LES NANCY CEDEX FRANCE

Abstract

Application of the hypertext paradigm to information retrieval requires 1) an automatic generation of hypertext links, 2) a compact graphical representation of the data. After a brief review of the family of neural algorithms required for deriving a compact and relevant representation of a documentary database, as well as links between synthetic "topics" and documents, we present a user interface based on these grounds. This representation is two-step : 1) a global topics map, 2) local topic axes, ranking both terms and documents according to the values of their "centrality index". A prototype, running in a Macintosh environment and implementing a basic version of this browser, is then described and commented.

Keywords

Hypertext, information retrieval, graphic user interface, neural networks, cluster analysis.

What representation of a documentary database are we aiming at ?

First, we will sketch the two major paradigms for designing user interfaces in our information retrieval field :

- With the query-answer paradigm, a dialog is supposed to take place between a user and a machine.

A major research trend has consisted in providing the system with a representation of the user. But the essential point is : how can the user have a relevant representation of what is the system, and of what is in the system ? Without it, he cannot adapt his action to his interlocutor. Considered from this point of view, most of the proposed systems do not take the issue into account (DACHELET 90).

- An important research line in recent years is related to the navigational paradigm (DANIEL-VATONNE 90) : the metaphor is that of navigation in the database, with powerful orientation tools at

Demander au Chat

Est-ce vraiment de l'IA ?



Demander au Chat

Est-ce vraiment de l'IA ?

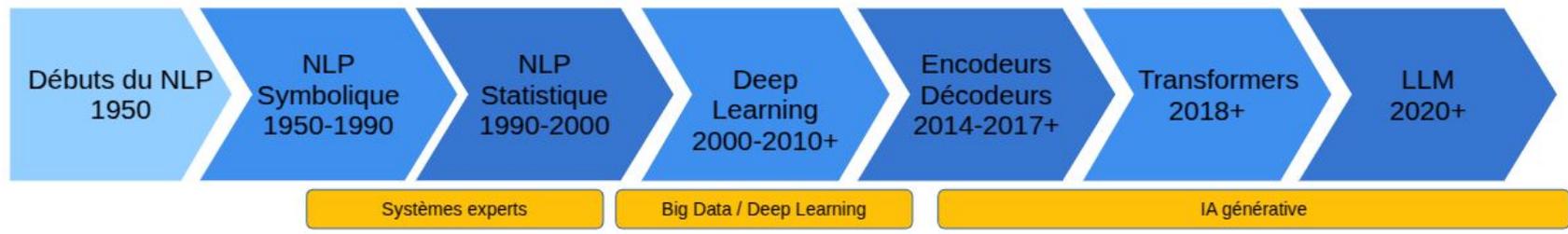
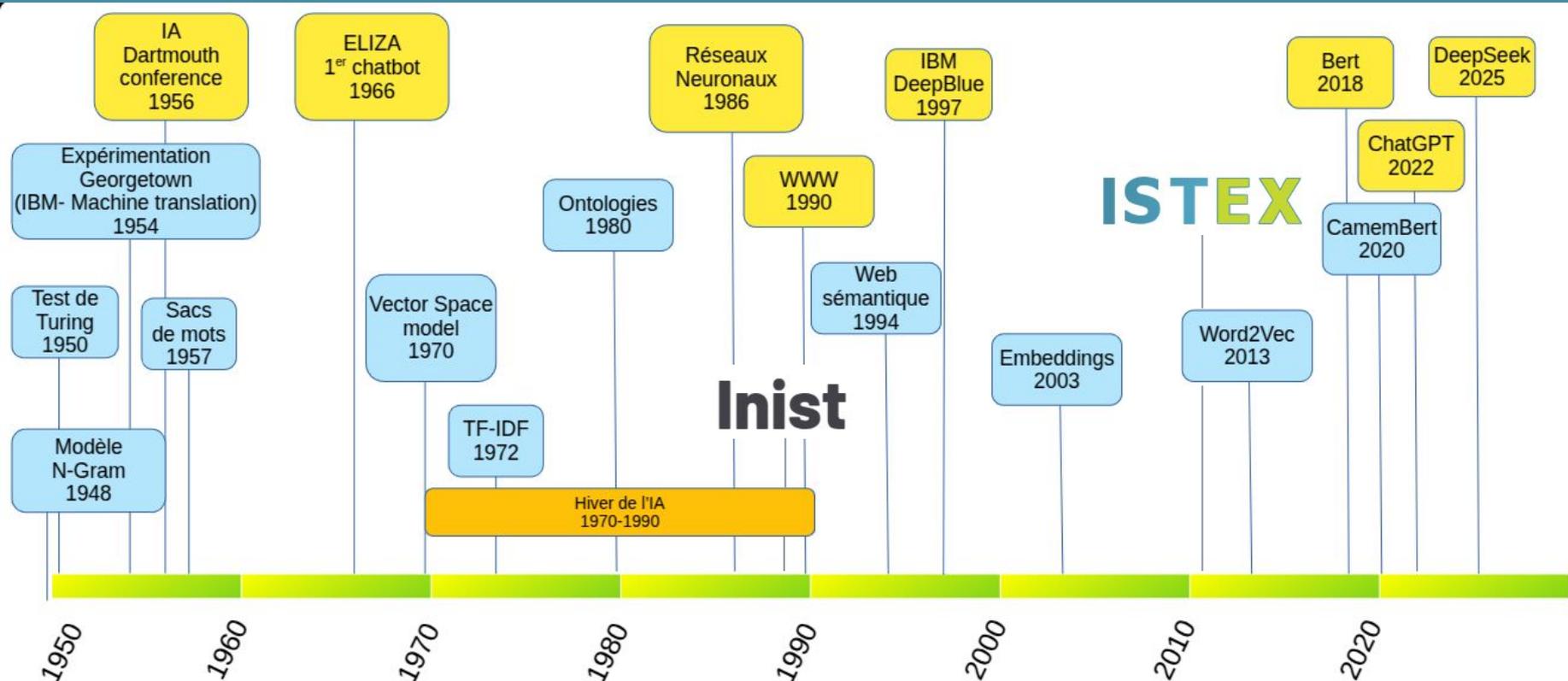


Oui, si on considère que la notion d'IA a vu le jour en 1955

A PROPOSAL FOR THE
DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE

J. McCarthy, Dartmouth College
M. L. Minsky, Harvard University
N. Rochester, I. B. M. Corporation
C. E. Shannon, Bell Telephone Laboratories

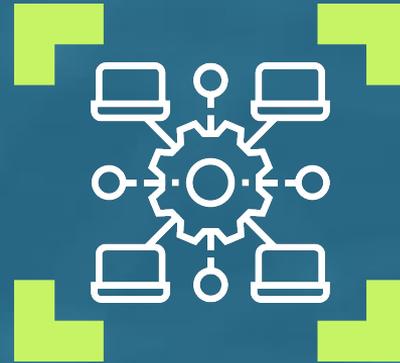
August 31, 1955



ISTEX

Infrastructure pour la fouille de textes

Istex



des textes & des services

48
bouquets

9 541
Revue

439 061
Monographies





28 219 752

documents

Soit une bibliothèque avec ~14 km de rayonnage

Un document, plusieurs versions du même texte



Plus de 300 000 000 de fichiers



Que faire de tous ces textes...

Analyse et fouille de textes



Contenu et accès
aux publications
intégrées **dans votre**
espace de travail

1. Accéder aux documents



Démarrer votre
projet de fouille
de texte en constituant
votre corpus à l'aide
des ressources Istex

2. Utiliser des corpus de documents



Les technologies
et les outils de
l'infrastructure Istex
au service de votre
projet de fouille de textes
sur **vos propres données**

3. Explorer et enrichir des corpus

Des services associés

Outils

ISTEX Search

Créez et téléchargez votre corpus scientifique

ISTEX TDM

Les **services Istex** pour la fouille de textes

ISTEX Lodex

Analyser vos propres données

Ressources

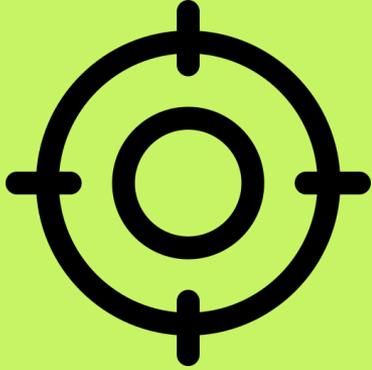
ISTEX Corpus

Les corpus Istex pour vos projets de re

ISTEX Loterre

Des terminologies en partage

Focus sur ISTEX TDM



42

~~Agents IA~~

Web Services

pour la fouille de textes.

textSummarize - Résumé automatique d'un article scientifique

Traitement

Introduction In bibliographic databases, affiliations of authors are of paramount importance. Hence, they permit to the laboratories or institutes to get national and even international visibility, as well as they consequently provide authors with scientific caution. We cannot discuss the issue of affiliations without talking of "Shanghai ranking" which aims at evaluating universities. Our purpose here is not to feed up the controversy (Van Raan 2005) (Liu, Cheng and Liu 2005), but to point out that the management of affiliation plays an important role in the calculation of universities "performance". Moed (Moed 2005) reports some problems with author's names and also institutions: "Authors from the same institution, or even from the same department, may not indicate their institutional affiliations in the same way". Depending on the country, it is not always clear how to name a laboratory with respect to its supervisory authorities. ... (Efficient supervised and semi-supervised approaches for affiliations disambiguation, Pascal Cuxac, Jean-Charles Lamirel, Valérie Bonvallot)



In bibliographic databases, affiliations of authors are of paramount importance. The affiliation is also important information to disambiguate author names in databases. As the amount of available information increases, problem caused by misspelling, spelling difference, and name or affiliation change also become worse. This paper proposes an approach based on Bayes learning method and overlapping clustering. It is structured as follows: section 1 summarizes related works, section 2 summarizes identifies problems, section 3 describes supervised learning method, and section 4 reports experiments. Section 5 concludes and discusses future work.

aiAbstract-check - Détection de résumé scientifique généré par IA

Traitement

The Conditional Lie-Bäcklund Symmetry (CLBS) method is introduced to analyze systems of evolution equations. It is demonstrated that the ability to reduce a system of evolution equations to a system of ordinary differential equations can be entirely described by the CLBS of the given system. As an application, a specific class of two-component nonlinear diffusion equations is examined. The governing system and its corresponding CLBS are determined. Consequently, exact solutions are derived based on polynomial, exponential, trigonometric, and mixed invariant subspaces through symmetry reductions."

⇒ {"isAiGenerated": true,"score": 1}

bibCheck - Contrôle de référence bibliographique

Traitement

2. Y. B. LINHART. L. CHAOUNI-BENABDALLAH, J.-M. PARRY & J. D. THOMPSON –
Selective herbivory of thyme chemotypes by a mollusk and a grasshopper

⇒

doi:10.3406/ecmed.1999.1879

status:found

3.<https://doi.org/10.1016/b978-0-323-90638-8.00002-3>

⇒

doi:10.1016/b978-0-323-90638-8.00002-3

status:found

4.Gerris Caucasicus, Primary Prevention of Cardiovascular Disease with a
Mediterranean Diet, 10.1056/nejmoa1200303

⇒

doi:10.1056/nejmoa1200303

status:retracted

5.Zohuri, B. (2019). A Comparison of Molten Salt Reactors to Light Water Reactors:
Pros and Cons. In Molten Salt Reactors and Thorium Energy (pp. 81-98). Woodhead
Publishing. <https://doi.org/10.1016/B978-0-08-102337-2.00006-9>

⇒

doi:

status:hallucinated

Teeft - Extraction de termes d'un texte via Teeft

Traitement

Mars Exploration Rover (MER) est une mission double de la NASA lancée en 2003 et composée de deux robots mobiles ayant pour objectif d'étudier la géologie de la planète Mars, en particulier le rôle joué par l'eau dans l'histoire de la planète. Les deux astromobiles ont été lancés au début de l'été 2003 et se sont posés en janvier 2004 sur deux sites martiens susceptibles d'avoir conservé des traces de l'action de l'eau dans leur sol. Chaque rover ou astromobile, piloté par un opérateur depuis la Terre, a alors entamé un périple en utilisant une batterie d'instruments embarqués pour analyser les roches les plus intéressantes : MER-A, rebaptisé Spirit, a atterri le 3 janvier 2004 dans le cratère Gusev, une dépression de 170 kilomètres de diamètre qui a peut-être accueilli un lac ; MER-B, renommé Opportunity, s'est posé le 24 janvier 2004 sur Meridiani Planum. Chaque rover pèse environ 185 kg et se déplace sur six roues mues par l'énergie électrique fournie par des panneaux solaires ...

⇒

```
"term": "deux robots",  
"frequency": 2,  
"specificity": 1"term": "panneaux solaires",  
"frequency": 2,  
"specificity": 1"term": "mars exploration rover mer",  
"frequency": 1,  
"specificity": 0.5"term": "mission double",  
"frequency": 1,  
"specificity": 0.5"term": "deux robots mobiles",  
"frequency": 1,  
"specificity": 0.5
```

TermSuite - Extraction de termes d'un corpus

Traitement

Ce web service ne traite pas des fichiers mais des corpus : ceci est un exemple de contenu d'un fichier d'un corpus

Microbiology and Nitrogen Cycle in the Benthic Sediments of a Glacial Oligotrophic Deep Andean Lake as Analog of Ancient Martian Lake-Beds . Potential benthic habitats of early Mars lakes, probably oligotrophic, could range from hydrothermal to cold sediments. Dynamic processes in the water column (such as turbidity or UV penetration) as well as in the benthic bed (temperature gradients, turbation, or sedimentation rate) contribute to supply nutrients to a potential microbial ecosystem ...



key: n: sediment

freq: 10

key: nn: proto-paratethys sea

freq: 9

key: a: glacial

freq: 7

key: n: mmes

freq: 7

key: a: tropical

freq: 7

dataHomogenise - Homogénéisation automatique de mots-clés

Traitement

1. [write , written , huge]
[immense , wrtie , example]
2. [... , ... , immense]
[huge , ... , ...]

⇒

1. [write , write , huge]
[huge , write , example]
2. [... , ... , immense]
[immense , ... , ...]

astroTag - Extraction d'entités nommées en astronomie

Traitement

V643 Orionis is a binary star system located in the Orion constellation, offering valuable insights into stellar evolution.

If these planets had isothermal atmospheres, the light curves obtained during those occultations would have been smooth. Hence the presence of spikes in the light curves demonstrates immediately that the atmospheres of Jupiter and Neptune are not isothermal over distances much smaller than a scale height

This did not occur accidentally since ϵ Carinae is one of the most luminous stars of the Galaxy. The environment of this extended object can be observed in very high angular (spatial) and spectral resolution simultaneously using the Hubble Space Telescope (HST) with its Space Telescope Imaging Spectrograph (STIS) on board (Kimble et al., 1988).

The purportedly fastest star in the Gaia catalogue, Gaia DR2 5932173855446728064 (Bromley et al. 2018; Marchetti et al. 2019), has been found to be spurious (Boubert et al. 2019).

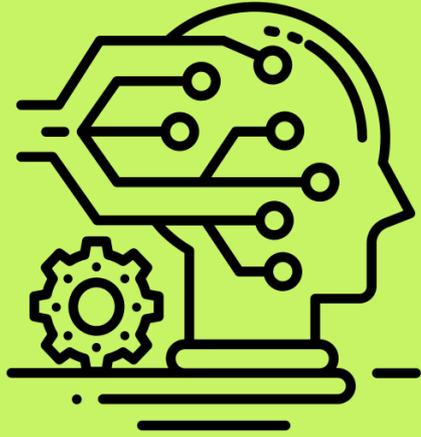
⇒ **Etoiles binaires (et pulsars):** V643 Orionis
Constellations: Orion

⇒ **Planete:** Jupiter, Neptune

⇒ **Objets artificiels:** Hubble Space Telescope, HST
Etoiles binaires (et pulsars): ϵ
Nebuleuse et region apparentes: Carinae

⇒ **Objets artificiels:** Gaia
Etoiles: Gaia DR2 5932173855446728064

Et aussi, pour la chimie, des maladies, des toponymes, espèces, etc.

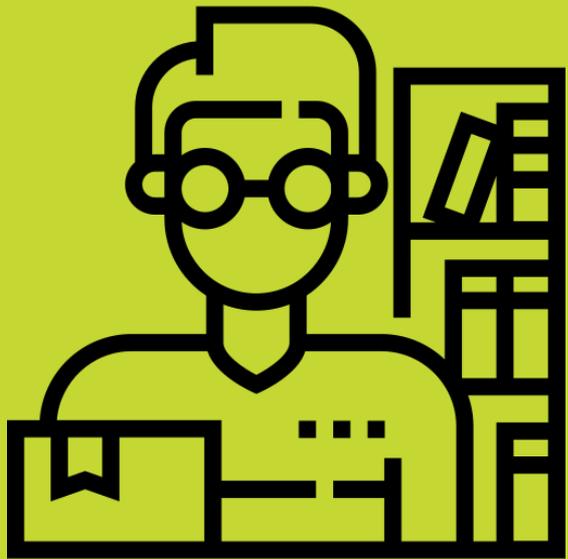


Mais comment s'en servir...



API

pour les personnes
qui connaissent la
programmation.



Lodex

pour les personnes
qui connaissent les
outils de **gestion de
données**



IA Factory

pour les personnes
qui connaissent les
outils bureautiques



Comment ça marche ?

Architecture technique

- ❖ Des **petits modèles spécifiques**
- ❖ Entraînement par GPU
- ❖ Exécution locale par CPU
- ❖ Code open source, Modèle en accès réservé

- ❖ Une architecture **extensible**



Quelle gouvernance ?



Nouvelle gouvernance en 2025

- Comité de pilotage
- Conseil Scientifique et Technique
- Comité opérationnel



Istex

Plusieurs équipes à votre service

Informations & Contact



Se tenir informé :

- Actualité et accès : <https://www.istex.fr/>



Chercher de l'aide :

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



Merci !

Des questions ?

