

Telescoop' Bubble



Genèse de Telescoop



Aussi,
Hatim alertait que suivre la
recherche sur les LLMs
devenait très difficile!

Panoram'IA

Panoram'IA est un **talk mensuel** traitant des **sujets d'IA** présenté par le **support utilisateur de l'IDRIS**.

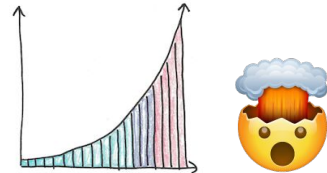
L'émission est composée de **Brèves** sur l'actualité IA, de **Revue De Papiers** scientifique et technique, de **Tests de l'IDRIS** et d'un **Paper Storm** des papiers les plus récents.

Vous pouvez voir et revoir les émissions **en live ou en replay** sur notre chaîne youtube : [Un oeil sur l'IDRIS](#) et sur notre playlist [Panoram'IA](#)

Prochaines émissions :

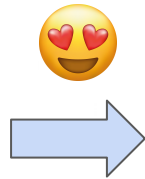
- 8 décembre 🎄
- à venir ..

Pour toutes questions ou demandes d'informations, vous pouvez nous contacter sur panoramia@idris.fr



~10000 papiers d'IA
publiés en septembre 2023
en croissance
exponentielle

Agrégateurs de publications scientifiques



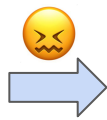
Mot clef : “Jean Zay Supercomputer”

Bloom: A 176b-parameter open-access multilingual language model
Flaubert: Unsupervised language model pre-training for french
Leading hadronic contribution to the muon magnetic moment from lattice QCD
What language model architecture and pretraining objective works best for zero-shot generalization?
Rita: a study on scaling up generative protein sequence models
Barthez: a skilled pretrained french sequence-to-sequence model
Un modèle Transformer Génératif Pré-entraîné pour le_ français (Generative Pre-trained Transformer in_ (French) We introduce a French adaptation from the well-known GPT model)
Drbert: A robust pre-trained model in french for biomedical and clinical domains
Transflower: probabilistic autoregressive dance generation with multimodal attention
KSPHPDDM and PCHPDDM: Extending PETSc with advanced Krylov methods and robust multilevel overlapping Schwarz preconditioners
One Versus all for deep Neural Network Incertitude (OVNNI) quantification
Denoising score-matching for uncertainty quantification in inverse problems
Improved constraints on reionisation from CMB observations: A parameterisation of the kSZ effect
High-resolution mining of the SARS-CoV-2 main protease conformational space: supercomputer-driven unsupervised adaptive sampling
...

~ 300 publications



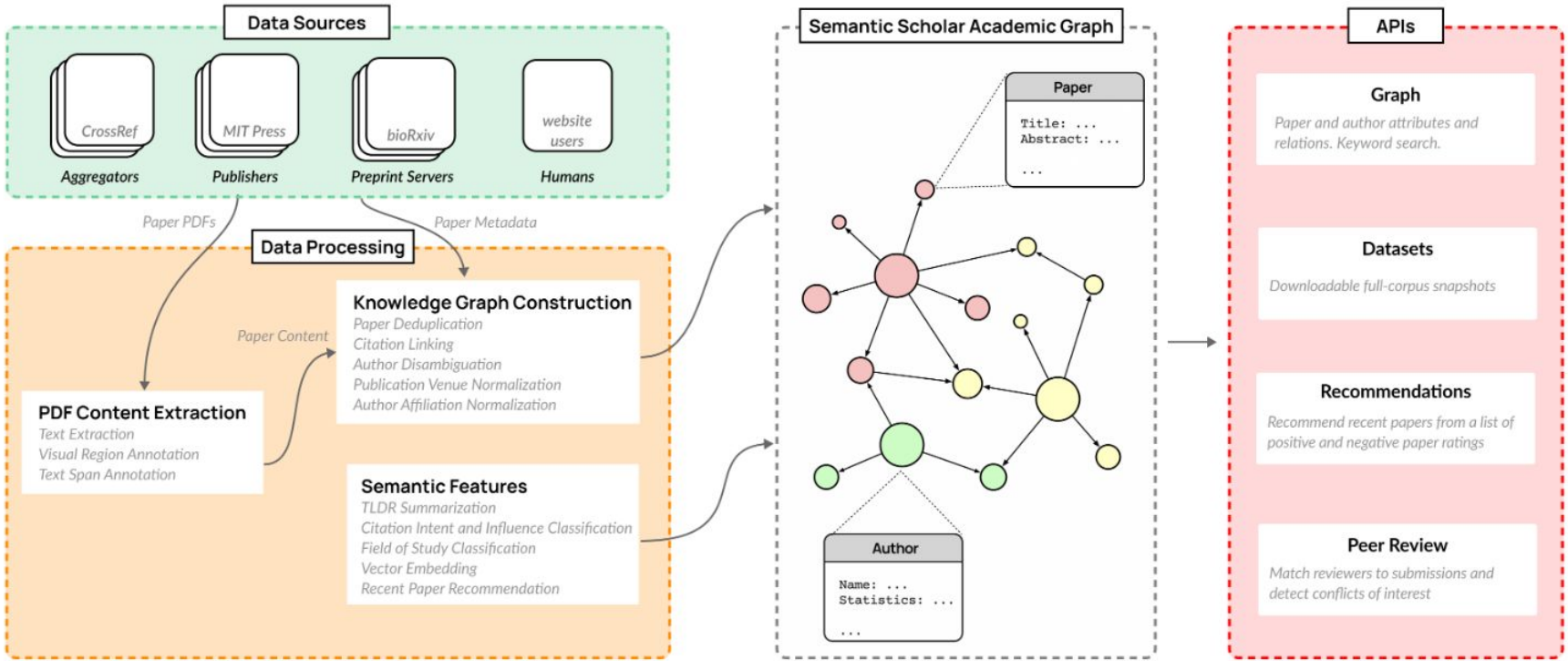
SEMANTIC SCHOLAR



7 publications ??

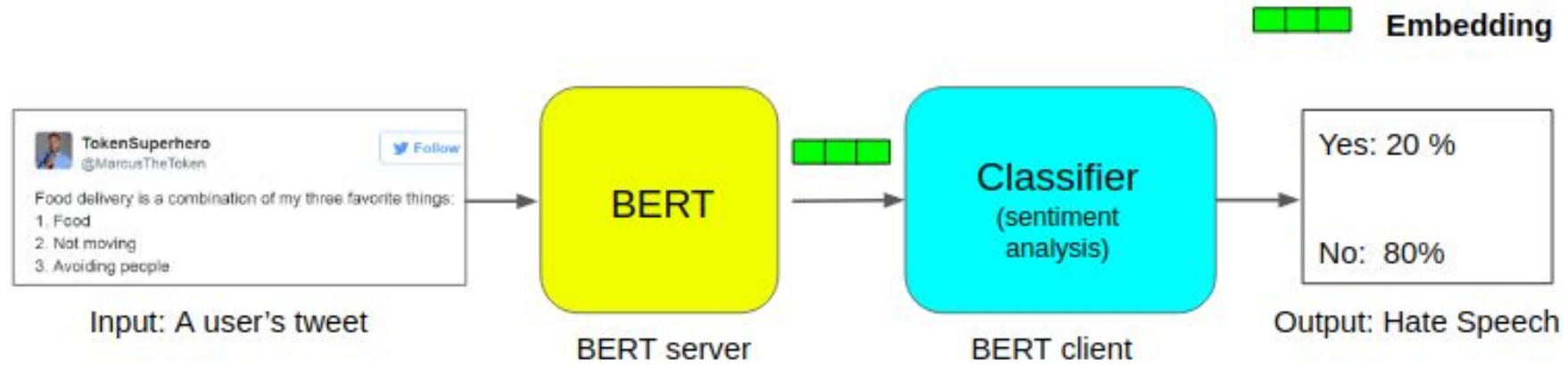
Semantic Scholar API

MAG
(arrêté en 2021) → S2AG

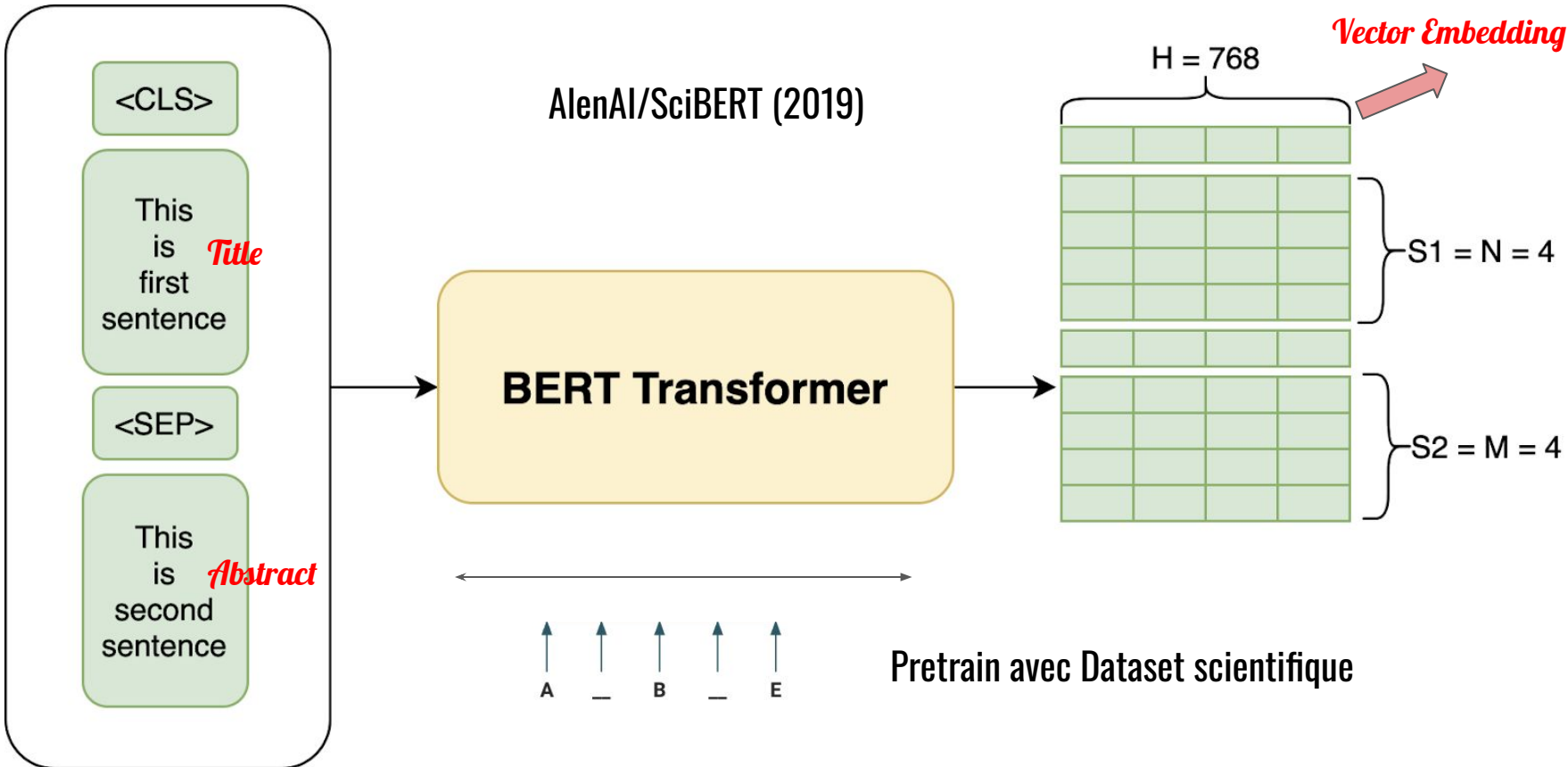


Vector Embedding - Representation learning

BERT qu'est ce que c'est ?!

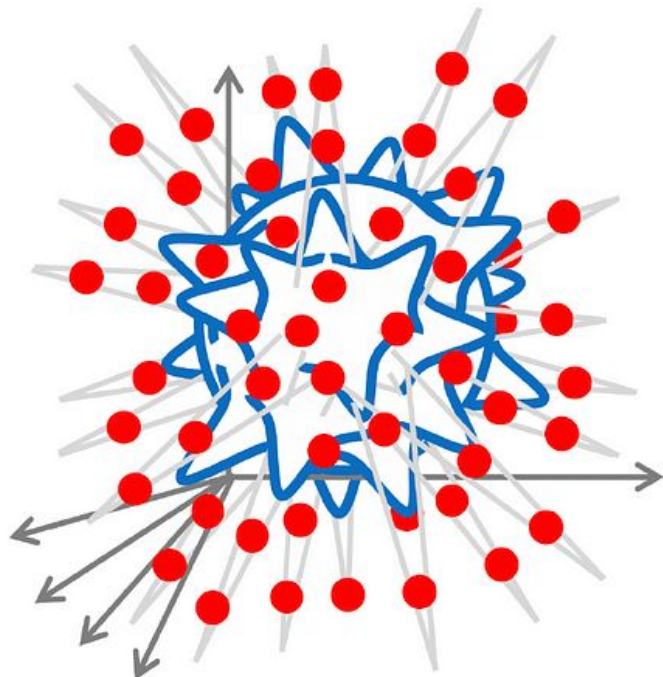


Vector Embedding - Representation learning



Vector Embedding - Representation learning

768 dimensions ???

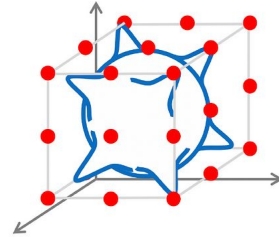
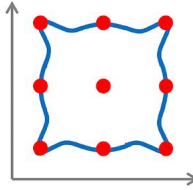


Hypercube

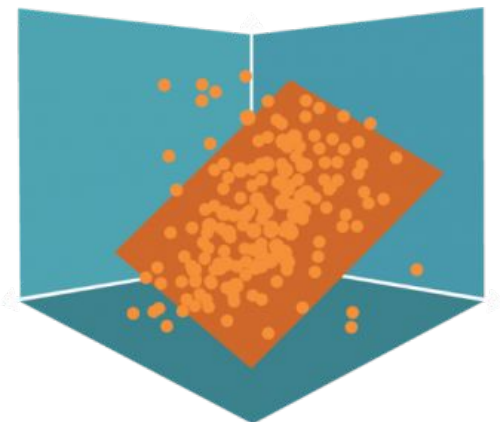
??



1D, 2D, 3D



Representation learning - PCA



3D Plane

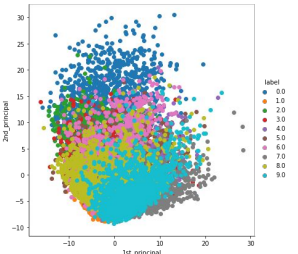
Dimensionality reduction
technique



2D Plane

Projection
Transformation linéaire

Representation - t-SNE (t-distributed Stochastic Neighbor Embedding)



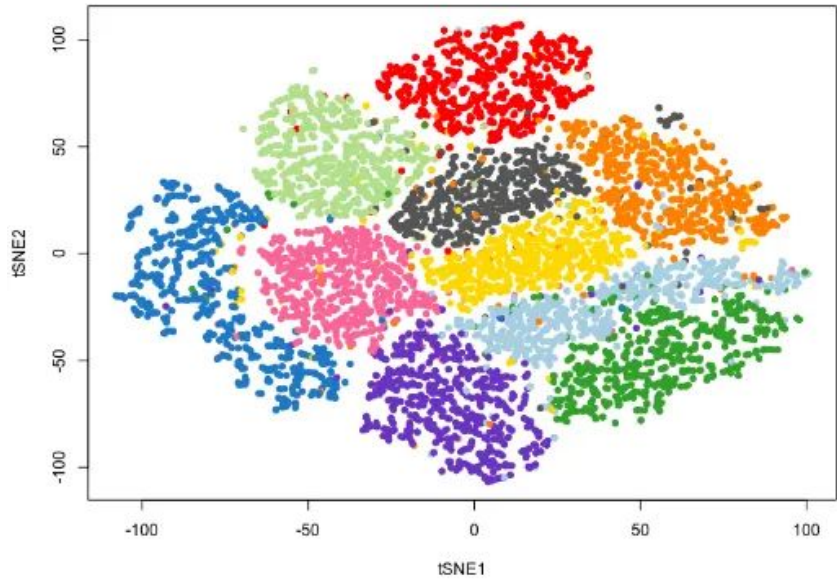
PCA



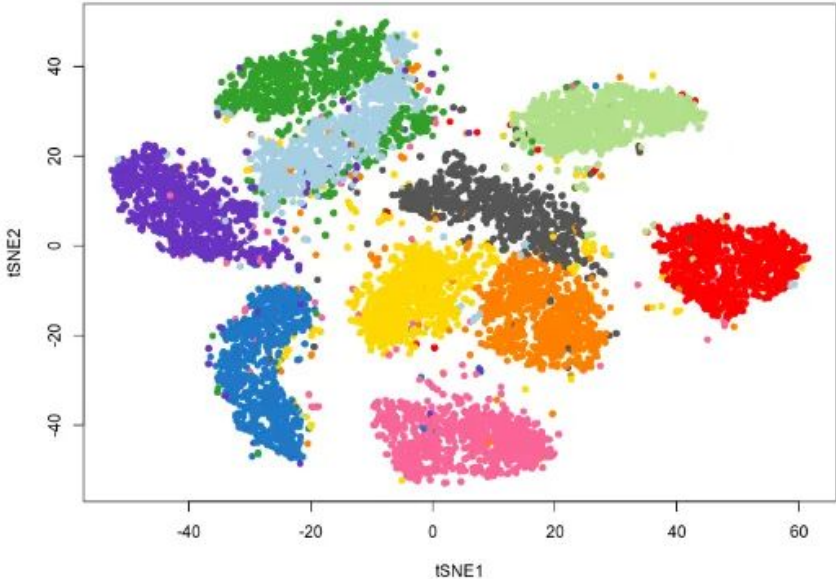
Seulement pour de la représentation 2D ou 3D

Transformation non-linéaire
et stochastique

tSNE MNIST: Perplexity = 5

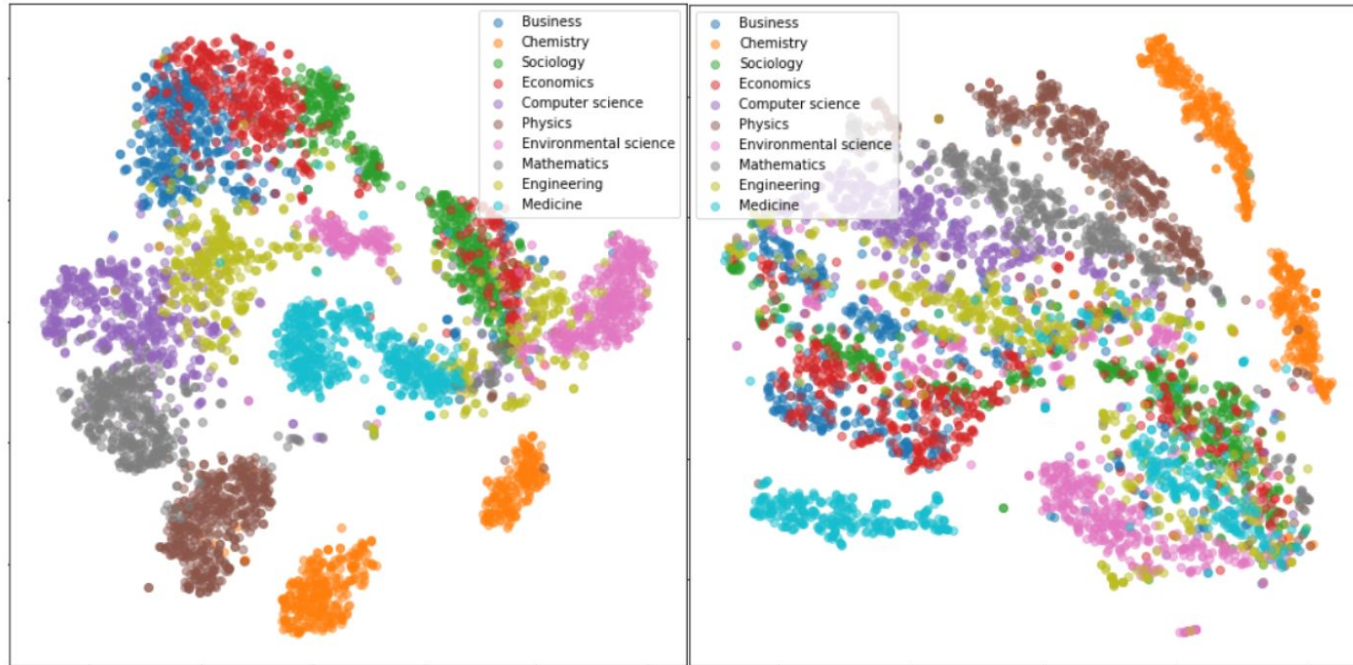


tSNE MNIST: Perplexity = 50



Vector Embedding - Representation learning

AllenAI/SPECTER (2020)

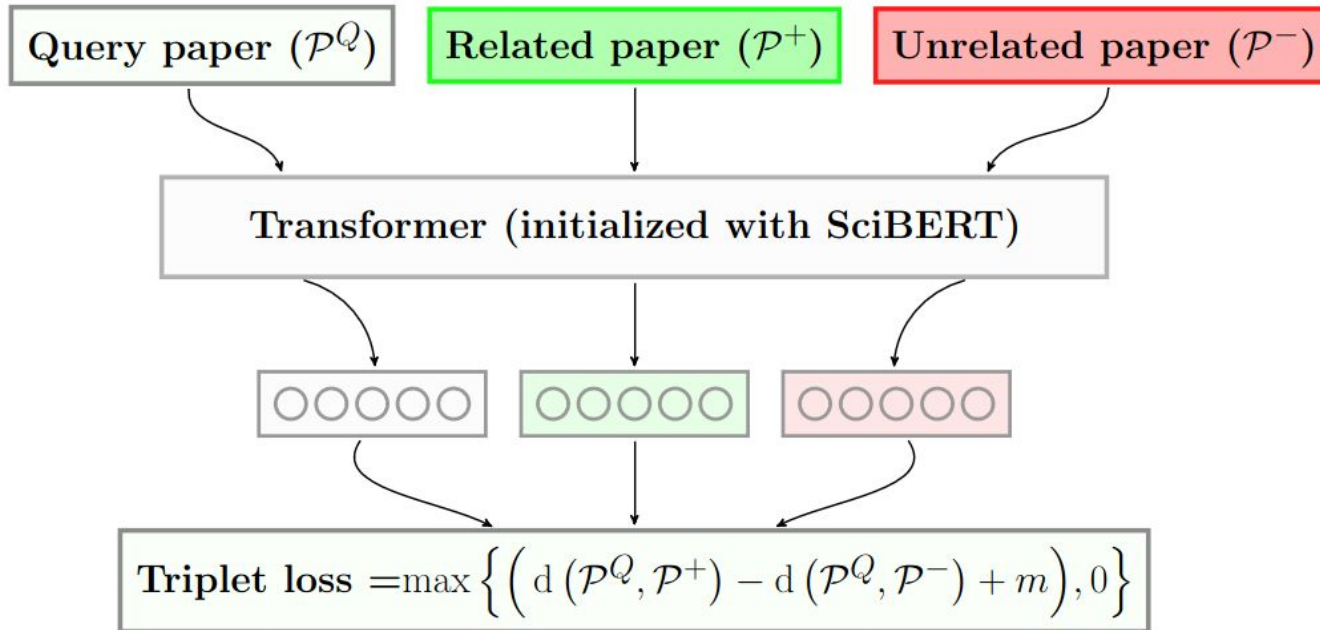


(a) SPECTER

(b) SciBERT

Vector Embedding - Representation learning

AllenAI/SPECTER (2020)



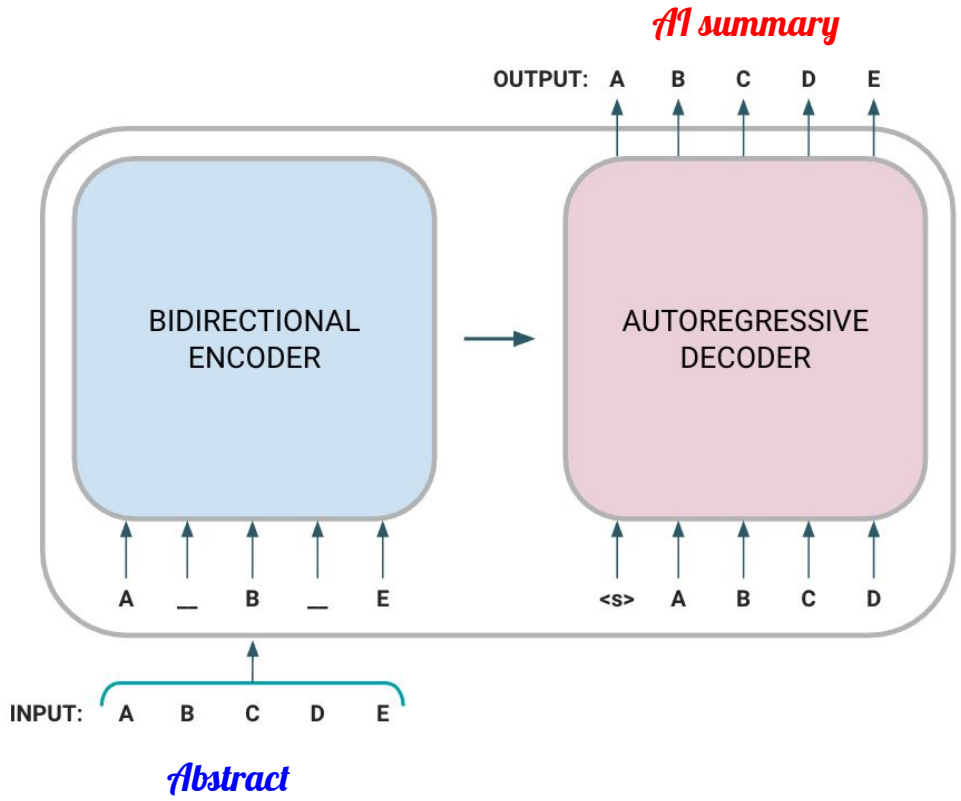
Vector Embedding - Representation learning

AlenAI/SPECTER2 (2023)

Field of study	SciRepEval (A)	SciDocs (B)	Increase Ratio (A/B)
Medicine	3,201,323	74,685	43
Computer Science	1,187,689	199,664	6
Biology	882,357	13,377	66
Chemistry	508,056	3,813	133
Psychology	492,071	22,590	22
Materials Science	271,865	7,681	35
Engineering	254,826	31,444	8
Mathematics	231,482	25,800	9
Physics	217,670	7,285	30
Business	217,585	5,450	40
Sociology	156,128	2,305	68
Political Science	154,388	1,032	150
Economics	123,357	2,705	46
Environmental Science	91,682	1,136	81
Art	89,527	206	435
Geography	83,688	1,491	56
Philosophy	61,996	151	411
Geology	51,103	640	80
History	46,430	159	292

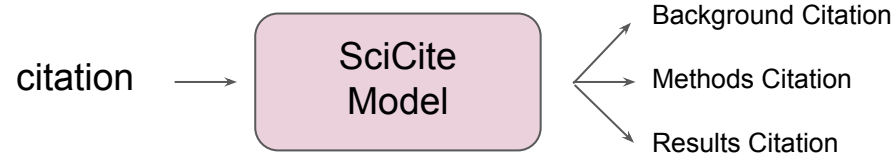
Table 9: Data domain distribution in SciRepEval for the training tasks and comparison with SciDocs. We group the unique documents in both the benchmarks by their MAG (Wang et al., 2020a) fields of study and present the counts in columns 2 and 3 and the absolute increase per field in column 4.

TLDR Summarization - BART



Citation Intent and Influence Classification

Citation Intent



Highly Influential Citation

Highly Influential ou pas ?

Heuristique



Highly Influential si et seulement si :

- (1) il n'y a pas d'auteurs en commun avec le papier citant,
- (2) La citation apparaît au moins à 3 endroits dans une phrase ou il n'y a pas d'autres citations, ou si la phrase citante contient des termes comme "build upon" "following" ou "inspired by", ou si la citation fait référence à des tables ou des figures.

Scholar's Hub

Stay Ahead of the Curve: Trending Papers

With the volume of new research published every day, we understand how hard it is to stay up-to-date with the latest discoveries. That's why we curated this collection of trending papers for you.

This list is curated based on the citations each paper receives per month and the page traffic growth on Semantic Scholar, along with other criteria. Currently only available for Biology, Computer Science, Medicine, Physics, and Psychology papers.

If you have any feedback or suggestions, please [contact us](#).

Last updated: October 23rd, 2023

Biology

Computer Science

Medicine

Physics

Psychology



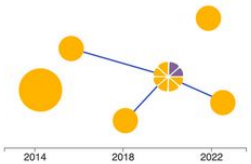
Semantic Scholar API Gallery

Semantic Scholar API

API Gallery

We're excited to have partners join in on our mission to accelerate scientific breakthroughs by building extraordinary tools on the Semantic Scholar APIs. Explore these use cases and get inspired for your next project!

If you'd like to add your project to the gallery, please fill out [this form](#).



Citrus Search

Get an overview of relevant articles in a research field with a single search.

Jonas Mayer, Maximilian Palmer



Sourcely

Find and summarize academic sources for students and academics writing their essays and papers.

Elman Mansimov

Better Together

Better Together

Input a corpus id or a query to find a list of similar papers with citation counts.

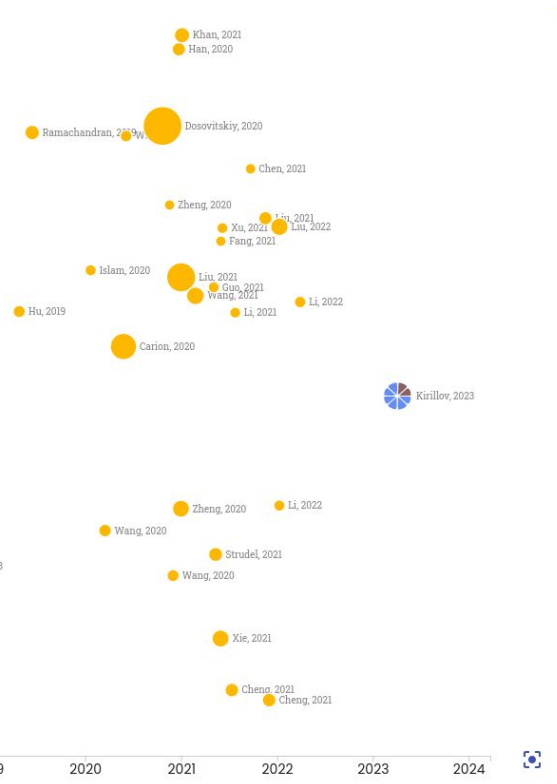
Kenneth Church

Citrus Search



Kirillov, 2023

[Feedback](#) [About](#)



All Results

- Search titles, authors, abstracts
- Seed papers**
- An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale**
18879 citations | 2020 | A. Dosovitskiy, Lucas Beyer et al.
 - Swin Transformer: Hierarchical Vision Transformer using Shifted Windows**
8952 citations | 2021 | Ze Liu, Yutong Lin et al.
 - End-to-End Object Detection with Transformers**
6736 citations | 2020 | Nicolas Carion, Francisco Massa et al.
 - Scene Parsing through ADE20K Dataset**
2079 citations | 2017 | Bolei Zhou, Hang Zhao et al.
 - Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions**
1940 citations | 2021 | Wenhai Wang, Enze Xie et al.
 - A ConvNet for the 2020s**
1853 citations | 2022 | Zhuang Liu, Hanzi Mao et al.
 - SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers**
1718 citations | 2021 | Enze Xie, Wenhai Wang et al.
 - Rethinking Semantic Segmentation from a Sequence-to-Sequence Perspective with Transformers**
1682 citations | 2020 | Sixiao Zheng, Jiachen Lu et al.



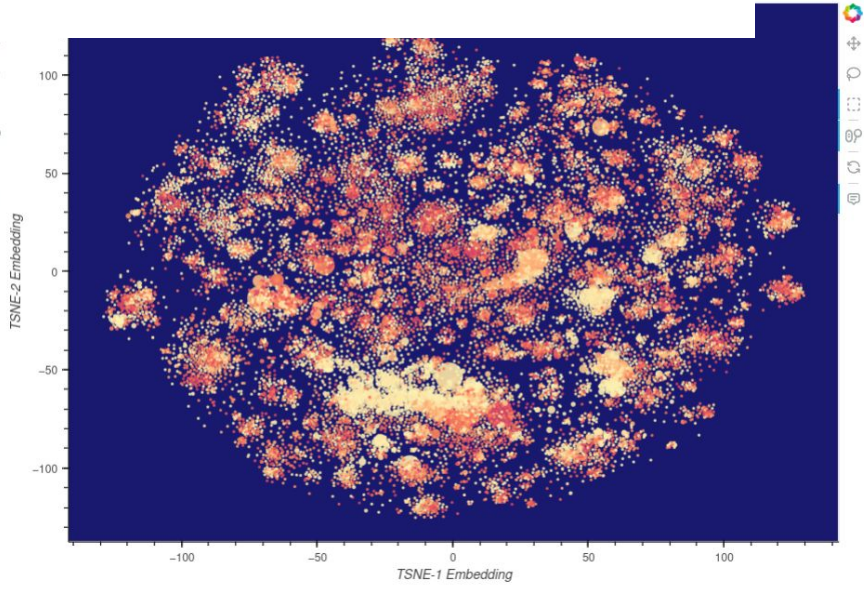
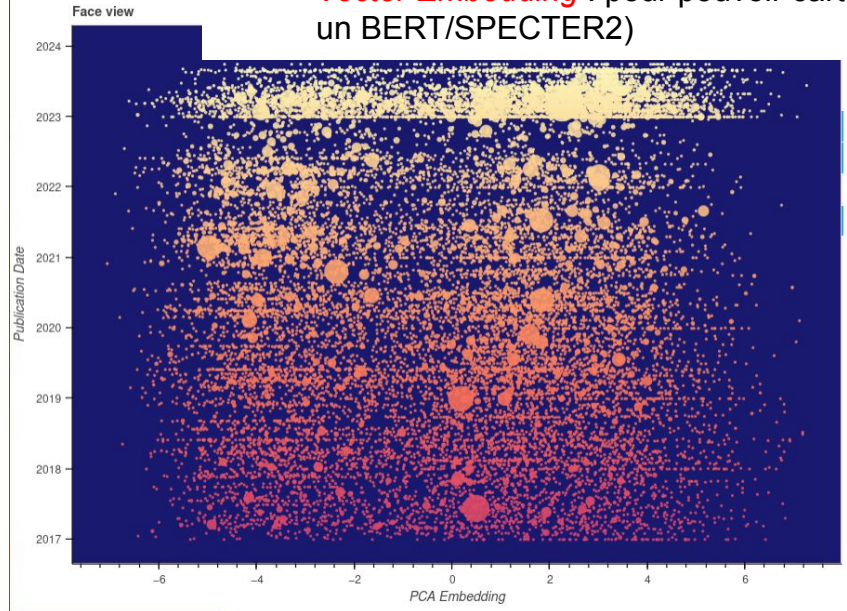
Telescope' Bubble

Circle size:

Select paper by title

Semantic Scholar features nécessaires :

- **Influential Citations Count** : pour mettre en valeur les papiers importants et influents
- **Vector Embedding** : pour pouvoir cartographier les papiers selon leur contenu (calculé par un BERT/SPECTER2)



Filter the table by selected papers

#	Title	Publication Date	Publication Year	Last Month Cit.	Cit. Acceleration	Influential Cit.	Citations
15	LLaMA: Open and Efficient Foundation Language Models	2023-02-27	2023	233	26	297	1407
20	GPT-4 Technical Report	2023-03-15	2023	186	52	113	1038
36	Llama 2: Open Foundation and Fine-Tuned Chat Models	2023-07-18	2023	129	44	36	237
78	Segment Anything	2023-04-05	2023	79	1	71	469
84	BLIP-2: Bootstrapping Language-Image Pre-training with Frozen Image Encoders and Large Language Models	2023-01-30	2023	77	18	117	455
85	Sparks of Artificial General Intelligence: Early experiments with GPT-4	2023-03-22	2023	76	-4	69	558
95	A Survey of Large Language Models	2023-03-31	2023	71	2	28	336

Évolutions possibles de Telescoop' Bubble

Aujourd'hui outil de visualisation et de représentation à destination des experts IA.

Evolutions possibles :

- Autres domaines d'expertise
- Plateforme plus interactive
- Classification fine automatique à partir de l'embedding
 - Clustering : k-means, DBSCAN
 - Machine learning : SVM, Neural Network

Possibilités dans un futur proche

Telescoop + PIE

IA Générative / LLM pour suivre la recherche :

- **Génération de newsletter** résumant les derniers papiers importants
- **Chatbot d'explication** de papier de recherche.
Big Challenge : valorisation de la réponse générée avec **citation des sources**