

Les usines à articles et les outils pour s'en prémunir avec Problematic Paper Screener

C.Labbé

Université Grenoble Alpes - LIG - équipe Sigma

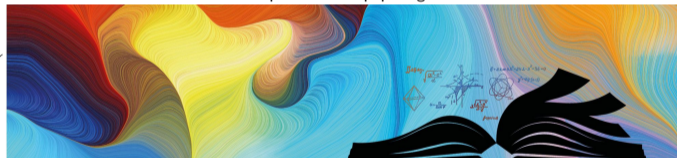


Table of Contents

- 1 Publications et Connaissances
 - La publication scientifique
 - Bibliométrie
- 2 Changement d'objectif : publier plus, citer plus
 - Générateur d'article: Grammaire Probabiliste Hors Contexte
 - Paraphraseurs
 - Citations furtives 🦹
- 3 Et la suite?
 - Vérifier – Nettoyer
 - Changement de paradigme

Table of Contents

- 1 Publications et Connaissances
 - La publication scientifique
 - Bibliométrie
- 2 Changement d'objectif : publier plus, citer plus
 - Générateur d'article: Grammaire Probabiliste Hors Contexte
 - Paraphraseurs
 - Citations furtives 🦹
- 3 Et la suite?
 - Vérifier – Nettoyer
 - Changement de paradigme

Exemple de publication 1944

[Avery et al., 1944]

**STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE
INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES****INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION
ISOLATED FROM PNEUMOCOCCUS TYPE III****By OSWALD T. AVERY, M.D., COLIN M. MACLEOD, M.D., AND
MACLYN McCARTY,* M.D.***(From the Hospital of The Rockefeller Institute for Medical Research)*

PLATE 1

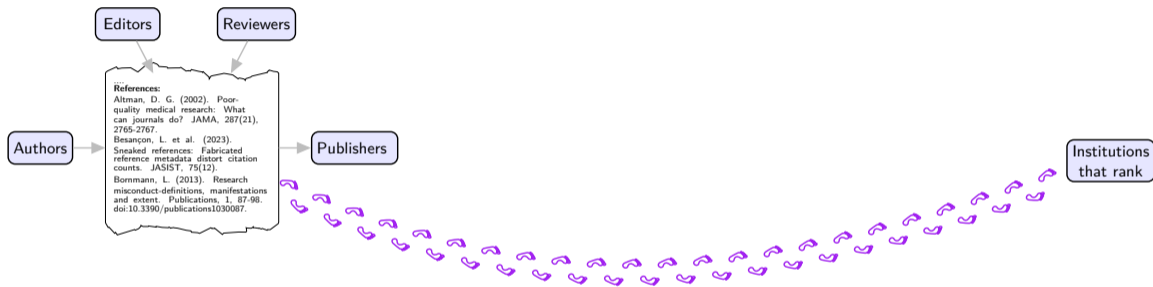
(Received for publication, November 1, 1943)

Biologists have long attempted by chemical means to induce in higher organisms predictable and specific changes which thereafter could be transmitted in series as hereditary characters. Among microorganisms the most striking example of inheritable and specific alterations in cell structure and function that can be experimentally induced and are reproducible under well defined and adequately controlled conditions is the transformation of specific types of Pneumococcus. This phenomenon was first described by Griffith (1) who succeeded in transforming an attenuated and non-encapsulated (2)

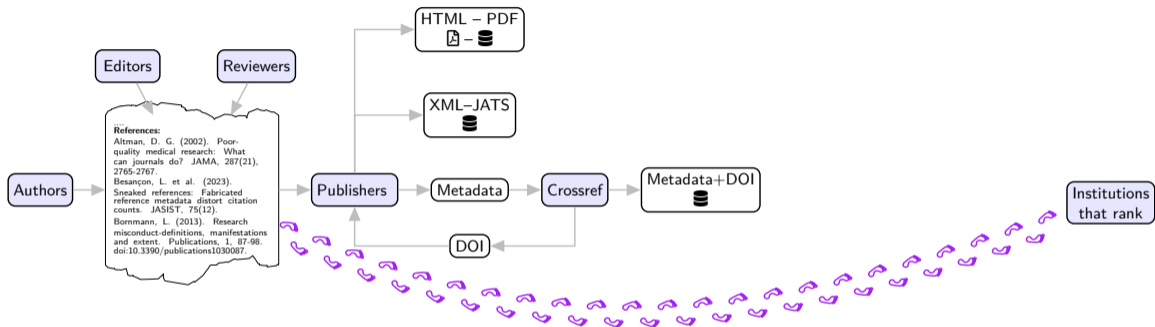
BIBLIOGRAPHY

1. Griffith, F., *J. Hyg., Cambridge, Eng.*, 1928, **27**, 113.
2. Neufeld, F., and Levinthal, W., *Z. Immunitätsforsch.*, 1928, **55**, 324.

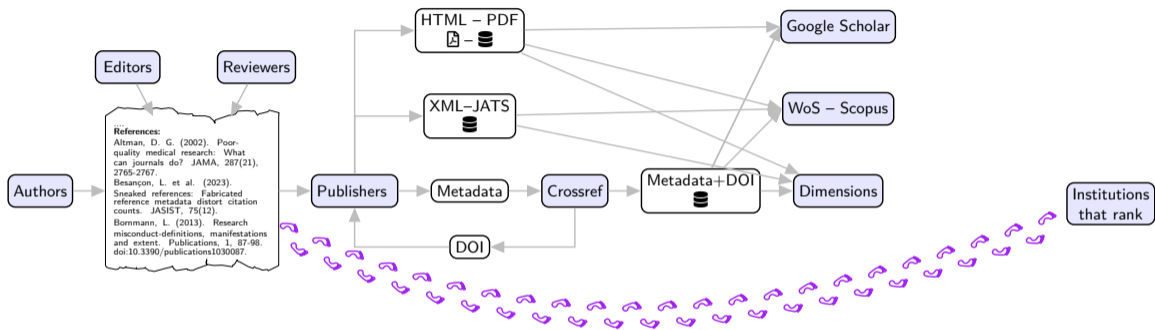
Information Systems for science



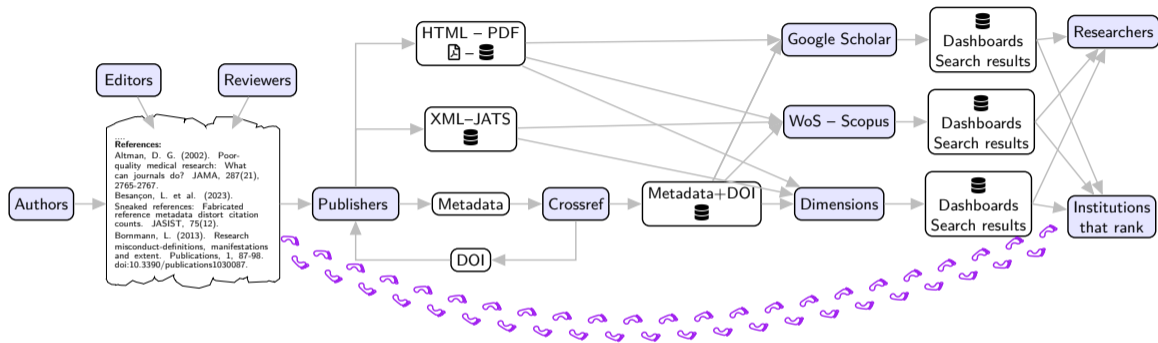
Information Systems for science



Information Systems for science



Information Systems for science



Scientific publications are at the heart of the system:

- Ideally: Knowledge diffusion.
- In practice: (Ac)Counting unit [Gingras, 2019].

ARWU: Academic Ranking of World Universities (Shanghai ranking) since 2003

ARWU methodology: from Wikipedia

Criterion	Indicator	Weighting	Source
Quality of education	(1) Alumni as Nobel laureates & Fields Medalists Weighting	10%	Official websites of Nobel Laureates & Fields Medalists
Quality of faculty	(2) Staff as Nobel Laureates & Fields Medalists.	20%	Official websites of Nobel Laureates & Fields Medalists
	(3) Highly cited researchers in 21 broad subject categories	20%	Thomson Reuters' survey of highly cited researchers (Clarivate)
Research output	(4) Papers in Nature and Science.	20%	Citation index
	(5) Papers indexed in SCI-ex and SSCI	20%	Citation index
Per capita performance	(6) Per capita academic performance of an institution	10%	–

You are the Dean:

What should you do to improve your ranking?

Prime à la publication – AEF info – Dépêche n°588621 – 04/10/2018

Exemples de primes à la publication d'un article en business school (en euros, rémunération nette, chiffres 2018) :

	Primes versées par 8 Business schools françaises classées par le <i>Financial Times</i>							
Étoiles des revues	BS 1	BS 2	BS 3	BS 4	BS 5	BS 6	BS 7	BS 8
*	1 500	1 000	500	350	-	1 000	1 500	800
**	3 000	2 000	2 000	750	2 500	2 000	4 500	1 500
***	4 500	5 000	10 000	4 000	5 000	3 000	7 500	5 000
****	6 000	10 000	15 000	12 000	10 000	5 000	12 000	10 000
Prime moyenne	3 750	4 500	6 875	4 275	4 375	2 750	6 375	4 325

Système d'Interrogation, de Gestion et d'Analyse des Publications Scientifiques

[Gingras and Khelifaoui, 2021]

Evaluation quantitative de la recherche des établissements de santé (CHU...) pour déterminer des dotations financières. Déployé en 2008.

Points rapportés par une publication

Journaux classés par *Impact Factor (IF)* :

Note	IF	points
A	top 10%, (< 10%)	8
B	10% < ... < 25%	6
C	25% < ... < 50%	4
D	50% < ... < 75%	3
E	> 75%	2
NC	in PubMed but not in WoS	1

Facteur multiplicatif fonction de la place de l'auteur

Les points de la publication sont multipliés par :

Facteur	Position de l'auteur
4	1er et dernier
3	2nd
2	3ième et avant dernier
1	Autres

Valeur du point (non fixé)

En 2017, 1 point vaut 648 € par an pour 4 ans.

1er (ou dernier) auteur pour une publication dans *Nature*

$$\Rightarrow 8 * 4 * 648 * 4 = 82\,944 \text{ €}$$

Citation Indexes for Science

[Garfield, 1955]

In this paper I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers.

Citation Indexes for Science

[Garfield, 1955]

In this paper I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers.

R. K. Merton letter to E. Garfield, August 3, 1974, RKMP 168.5 – reported by [Csiszar, 2020]

Watch out for goal displacement: Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did.

Citation Indexes for Science

[Garfield, 1955]

In this paper I propose a bibliographic system for science literature that can eliminate the uncritical citation of fraudulent, incomplete, or obsolete data by making it possible for the conscientious scholar to be aware of criticisms of earlier papers.

R. K. Merton letter to E. Garfield, August 3, 1974, RKMP 168.5 – reported by [Csiszar, 2020]

Watch out for goal displacement: Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did.

Computed yearly since 1975: *Impact Factor*

Year y : Average number of citations received that year to papers published by the journal over the last two years ($y - 1, y - 2$).

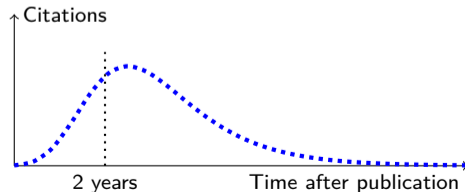


Table of Contents

- 1 Publications et Connaissances
 - La publication scientifique
 - Bibliométrie
- 2 Changement d'objectif : publier plus, citer plus
 - Générateur d'article: Grammaire Probabiliste Hors Contexte
 - Paraphraseurs
 - Citations furtives 🦹
- 3 Et la suite?
 - Vérifier – Nettoyer
 - Changement de paradigme

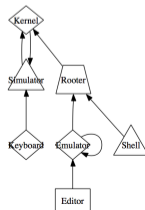
SCIgen ... maximize amusement, rather than coherence ... 2005 by J. Stribling, M. Krohn & D. Aguayo

Router: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interoperable.



The rest of this paper is organized as follows. For starters, we motivate the need for fiber-optic cables. We place our work in context with the prior work in this area. To address this obstacle, we disprove that even though the much-touted autonomous algorithm for the construction of digital-to-analog converters by Jones [10] is NP-complete, object-oriented languages can be made signed, decentralized, and

REFERENCES

- [1] S. Abiteboul, Y. Huang and V. Ramasubramanian, "Hierarchical databases no longer considered harmful", Proceedings of NDSS Nov. 2005, pp. 22-28.
- [2] O. Dahl, D. Johnson and R. Turing, "A. Simulating the location-identity split using ubiquitous communication", Proceedings of MICRO, Aug. 2006, pp.34-38.

SCIgen ... maximize amusement, rather than coherence ... 2005 by J. Stribling, M. Krohn & D. Aguayo

Router: A Methodology for the Typical Unification of Access Points and Redundancy

Jeremy Stribling, Daniel Aguayo and Maxwell Krohn

ABSTRACT

Many physicists would agree that, had it not been for congestion control, the evaluation of web browsers might never have occurred. In fact, few hackers worldwide would disagree with the essential unification of voice-over-IP and public-private key pair. In order to solve this riddle, we confirm that SMPs can be made stochastic, cacheable, and interoperable.



The rest of this paper is organized as follows. For starters, we motivate the need for fiber-optic cables. We place our work in context with the prior work in this area. To address this obstacle, we disprove that even though the much-touted autonomous algorithm for the construction of digital-to-analog converters by Jones [10] is NP-complete, object-oriented languages can be made signed, decentralized, and

REFERENCES

- [1] S. Abiteboul, Y. Huang and V. Ramasubramanian, "Hierarchical

La famille Scigen :

L'ancêtre POMO (*Postmodernism Generator*) [[Bulhak, 1996](#)]

SCIgen, Mathgen, scigen-physics, SBIR Proposal Generator...

Editor

location-identity split using ubiquitous communication ,

Proceedings of MICRO, Aug. 2006, pp.34-38.

<https://arxiv.org/pdf/2210.15112v1>

Let $s \leq X$ be arbitrary.

Definition 2.8. Assume every positive definite, elliptic, solvable modulus is almost negative. A Kronecker system is a **curve** if it is connected and quasi-solvable.

Definition 2.9. Let us assume there exists a globally convex solvable, quasi-everywhere Hilbert–Markov curve. A holomorphic, ultra-Steiner topos is a **homomorphism** if it is connected.

Proposition 2.10. $\tilde{\mathbf{k}} \geq D_\Omega$.

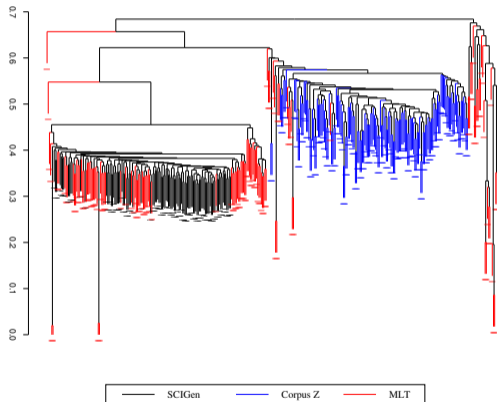
Proof. This proof can be omitted on a first reading. Let \bar{H} be a multiply singular, hyper-differentiable subalgebra. Trivially, if $Z_{f,\tau}$ is co-totally Tate–Brouwer then $y = 1$. So $-1^9 \sim t(\bar{q} \cap 1, -\pi)$. Now there exists a discretely universal hull. Hence every multiply normal domain is right-trivially universal. Next, if $|S| < |j|$ then ψ is dominated by f . Thus if V is ultra-everywhere hyperbolic and symmetric then $H = \sqrt{2}$. On the other hand, there exists a super-continuously smooth and elliptic triangle. By countability,

$$\begin{aligned} \overline{\Delta}'' &> \left\{ 2\sqrt{2}: \exp(0|\psi|) \neq \delta''^{-1}(Z) \cap \exp^{-1}(\mathscr{W}(\Theta)^{-2}) \right\} \\ &\geq \int_0^e p\left(\frac{1}{\bar{\mu}}, \dots, -\infty\right) db - a'\left(c, \dots, \frac{1}{d'}\right) \\ &\rightarrow \bigotimes_{-\infty} Z^{-1}(-\infty) \end{aligned}$$

Détection automatique – Préventif

[Labbé and Labbé, 2013]

Clustering Hiérarchique



Classification (plus proches voisins)

Distance Intertextuelle :

$\Delta_{(a,b)} = \delta$ proportion of different word tokens in the two texts.

t text à tester. :

$$\bullet \delta_t^{Fake} = \min_{f \in SCIGen} \Delta_{(t,f)}$$

Si ($\delta_t^{Fake} < \delta_{threshold}$) then

Génération SCIGen à considérer
(risk $< 10^{-5}$).

Sinon

Origine Non-SCIGen à considérer.

Détection automatique – Curatif

[Cabanac and Labbé, 2021]

Phrase search :

Many SCI_PEOPLE would agree that, had it not been for SCI_GENERIC_NOUN, ...
In recent years, much research has been devoted to the SCI_ACT; ...
SCI_THING_MOD and SCI_THING_MOD, while SCI_ADJ in theory, have not until ...
The SCI_ACT has SCI_VERBEDSCI_THING_MOD, and current trends ...
The implications of SCI_BUZZWORD_ADJ SCI_BUZZWORD_NOUN have ...

Exercice :

Utiliser Google Scholar avec la requête
"would agree that, had it not been for".

Problematic Paper Screener PPS – G. Cabanac – IRIT :

Collecter les papiers problématiques avec les résultats de requêtes (*fingerprints*) sur un moteur de recherche académique (*Dimensions.ai*)

Tortured Phrases – Paraphraseurs

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] – GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, **contrasted with the overall mumble**, has bpreparationstrated to be moderately effective. The **exactness rate** for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

Tortured Phrases – Paraphraseurs

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] – GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, **contrasted with the overall mumble**, has bpreparationstrated to be moderately effective. The **exactness rate** for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have showed that sector based tweet classification is more successful **compared to general tweets**. **Accuracy rates** for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.

Tortured Phrases – Paraphraseurs

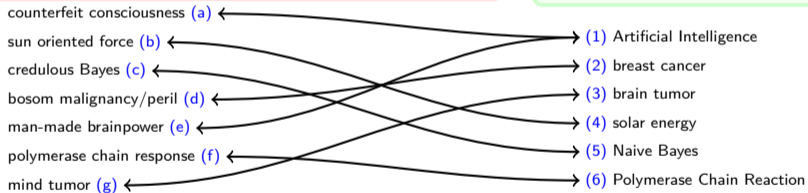
[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] – GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, **contrasted with the overall mumble**, has bpreparationstrated to be moderately effective. The **exactness rate** for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have showed that sector based tweet classification is more successful **compared to general tweets**. **Accuracy rates** for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.



Tortured Phrases – Paraphraseurs

[Cabanac et al., 2021]

Questionable text in case 6, rewritten from [13] – GPT detector score: 59.20%

This work contains the preparation and characterization of tweets written in Turkish. Tweet information sets a vector of **four different offices**, contrasted and the outcomes and the installed model and grouping support vector machine and the **arbitrary timberland arrangement** in Word. Area-based tweet arrangement of, **contrasted with the overall mumble**, has been prepared to be moderately effective. The **exactness rate** for the financial area 89.97 percent, soccer 84.02 percent, 73.86 percent for correspondence, it has been made 74.60 percent of the absolute of the 63.68 percent for retail.

Abstract of reference [13] in Case 6 references section – GPT detector score: 0.02%

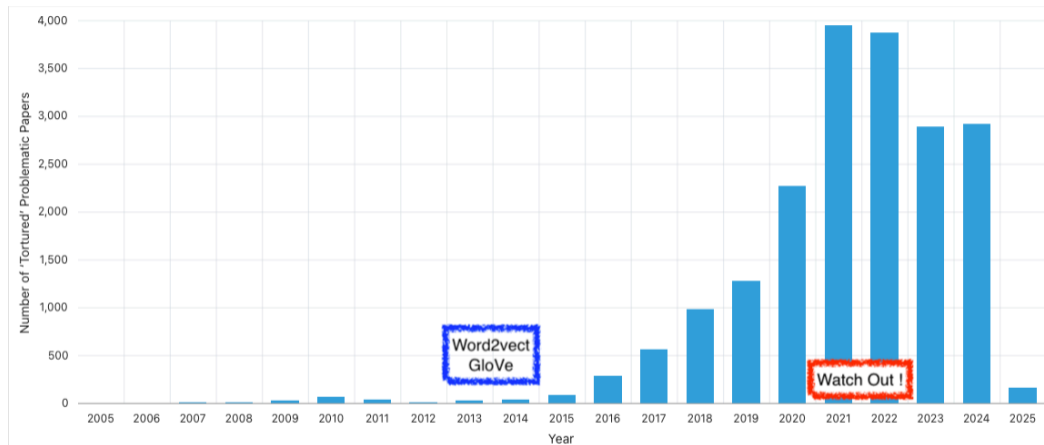
This work includes processing and classification of tweets which are written in Turkish language. **Four different sector** tweet datasets are vectorized with Word Embedding model and classified with Support Vector Machine and **Random Forests classifiers** and results have been compared. We have showed that sector based tweet classification is more successful **compared to general tweets**. **Accuracy rates** for Banking sector is 89.97%, for Football 84.02%, for Telecom 73.86%, for Retail 63.68% and for overall 74.60% have been achieved.



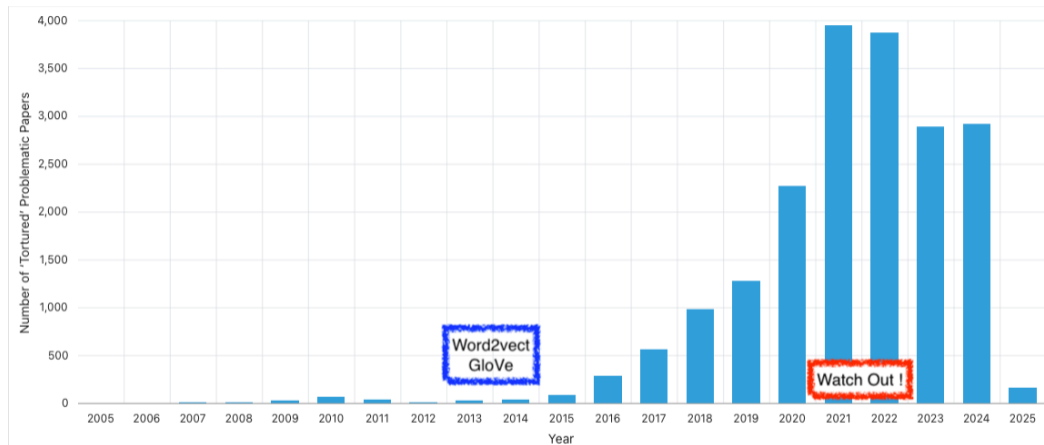
Détection automatique :

Les phrases torturées comme *fingerpint*

PPS: tortured phrases as fingerprints <https://www.irit.fr/Guillaume.Cabanac/problematic-paper-screener>



PPS: tortured phrases as fingerprints <https://www.irit.fr/Guillaume.Cabanac/problematic-paper-screener>



- Prévalence papiers torturés (2021) : 3 / 10 000
- 68% affiliation en Inde — 9% affiliation en Chine

Semantic Scholar: <https://openscilm.allen.ai> (mai 2025)

Related work on bosom malignancy

Answer

Share Disclaimer

Bosom cancer is the most common and hazardous disease in females and the second main explanation of disease-related passings in females, after lung cancer ([Hausalmaal. 2021¹](#)). According to the International Agency for Research on Cancer (IARC), which is part of the World Health Organization (WHO), the number of passings considered by cancer in the year 2012 was around 8.2 million ([Iqbal et al. 2022¹](#)). The number of new

La première référence est un "*Lazy Survey*" avec des phrases et des acronymes torturées :

- ... utilizing convolutional neural systems (CNNs)...
- ... The ejection of bogus positive (FP)...
- ... bosom tumor (BC) ...
- ... Bolster vector machine ...

Sneaked references – 👻 – [Besançon et al., 2024][Besançon et al., 2025]

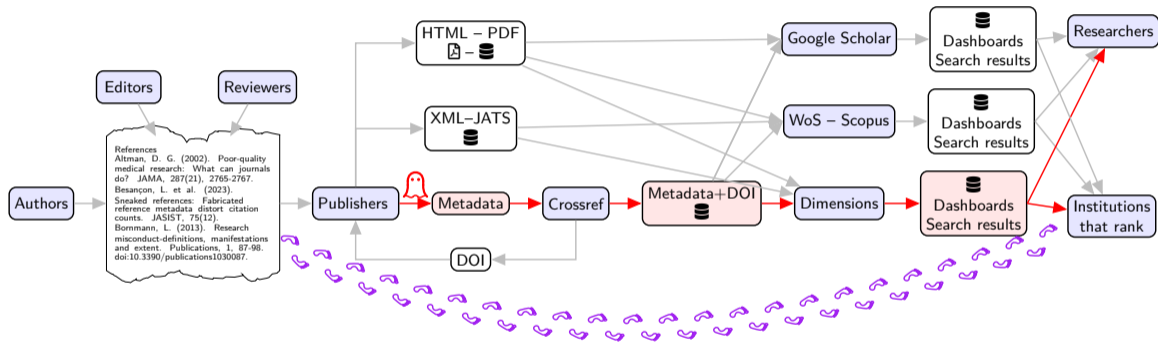


Table of Contents

- 1 Publications et Connaissances
 - La publication scientifique
 - Bibliométrie
- 2 Changement d'objectif : publier plus, citer plus
 - Générateur d'article: Grammaire Probabiliste Hors Contexte
 - Paraphraseurs
 - Citations furtives 🦹
- 3 Et la suite?
 - Vérifier – Nettoyer
 - Changement de paradigme

Tools – Pre- and Post- Publication Review



How, when and why science does science fail to correct itself?

Study mechanisms that hinder the correction of fraud and error.

- History and sociology of science, strong focus on nano-biology,
- *Journals and publishers*
- Build tools to analyse scientific literature :
 - Détection de phrases torturées (Alexandre Clause) [Clause et al., 2025]
 - Impact d'une rétractation sur un papier citant (Yağmur Öztürk) [Ozturk et al., 2024]
 - Détection de citations erronées (Qinyue Liu) [Liu et al., 2024]

Post Publication Peer Review – PubPeer – <https://pubpeer.com/>



- PubPeer (2012) déclencheur de "grosses" affaires françaises.
- Site web permettant de commenter anonymement ou non... (élément de succès et de controverse).
- Les commentaires sont modérés et ne doivent concerner que des éléments vérifiables.
- En 2024 PubPeer reçoit le prix "research integrity" de la "Einstein Foundation (Germany)"

DORA [URL]

...

Ces recommandations s'articulent autour d'un certain nombre de sujets :

- la nécessité de mettre un terme à l'utilisation d'indicateurs basés sur les revues, comme les facteurs d'impact, dans le financement, les nominations et les promotions ;

...

Pour les établissements

4. Afficher explicitement les critères utilisés dans les décisions de recrutement, de titularisation et de promotion, en soulignant clairement, surtout pour les chercheurs débutants, que le contenu scientifique d'un article est beaucoup plus important que les indicateurs de publication ou l'image de marque de la revue dans laquelle il a été publié.



...

Dutch national research foundation h-index ban [URL]


4. Why are the h-index and journal impact Factor not allowed to be mentioned?



Responsible assessment of research is essential for NWO. As a funder, we want to responsibly assess and compare individual researchers. Citing the h-index or journal impact factor are not allowed, as they do not give a reliable picture of the quality and contribution of the individual researcher.

Extensive research has demonstrated this (see notes). This conclusion is also supported by [a letter](#)  from the [AWTI](#)  (Advisory Council for Science, Technology and Innovation), in which they examined how to determine the quality of science. This letter therefore reinforces our policy not to use these indicators.



Waltman, L., & van Eck, N. J. (2012). *The inconsistency of the h-index*. *Journal of the American Society for Information Science and Technology*, 63(2), 406-415. <https://doi.org/10.1002/asi.21678> 



Halt the h-index. (2021, mei 19). <https://www.leidenmatrics.nl/articles/halt-the-h-index> 



Section 37

Economie et gestion

Motion

Communiqué de la Section 37 du CoNRS

Dans le contexte national et international de mise en œuvre des politiques publiques de la science ouverte et en accord avec l'engagement de la recherche (DORA : <https://sfdora.org/read/fr/>) signé et mis en œuvre par le CNRS, la décision a été prise de renoncer à toute forme de classement de revues en économie et en gestion.

Claude DIEBOLT
Président de la Section 37

Motion adoptée le 4/12/2020

21 votants : 21 oui, 0 abstention, 0 non



ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

15 novembre 2023

Avis du Comité Évaluation et science ouverte (CoÉSO) de l'Académie des sciences sur le score SIGAPS

Recommandations

Le CoÉSO de l'Académie des sciences appelle :

- Les responsables universitaires (présidents d'université, doyens de faculté, présidents de CNU, etc.) **à renoncer à l'utilisation du score SIGAPS pour évaluer les dossiers des candidats et candidates à des recrutements ou des promotions hospitalo-universitaires au profit de leur évaluation qualitative ;**
- Dans le cadre de l'évaluation de la production scientifique globale d'un centre hospitalier, pour laquelle le recours à un score quantitatif peut s'avérer utile, à une **modification de la méthode de calcul du score SIGAPS pour privilégier la qualité du contenu des publications**. Une solution pourrait **provenir de la prise en compte de l'impact effectif de chaque article (citations ultérieures dans la littérature scientifique) dans le calcul du score ;**

R. K. Merton letter to E. Garfield, August 3, 1974, RKMP 168.5 – reported by [Csiszar, 2020]

Watch out for goal displacement: Whenever an indicator comes to be used in the reward system of an organization or institutional domain, there develop tendencies to manipulate the indicator so that it no longer indicates what it once did.

References

Thanks to: Barhomi, A. – Bordignon, F. – Byrne, J. A. – Cabanac, G. – Capes-Davis, A. – Clause, A. – Favier, B. – Gautier, T. – Labbé, D. – Lay, P. – Lentschat, M. – Magazinov, A. – Nguyen, M. T. – Öztürk, Y. – Park, Y. – Portet, F. – Liu, Q. – Te, S. – West, R.



Avery, O. T., MacLeod, C. M., and McCarty, M. (1944).

Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III.

Journal of Experimental Medicine, 79(2):137–158.



Besaçon, L., Cabanac, G., Labbé, C., and Magazinov, A. (2024).

Sneaked references: Fabricated reference metadata distort citation counts.

Journal of the Association for Information Science and Technology, 75(12):1368–1379.



Besaçon, L., Cabanac, G., Labbé, C., Magazinov, A., di Scala, J., Tkaczyk, D., and

Weber-Boer, K. (2025).

Detection of metadata manipulations: Finding sneaked references in the scholarly literature.

CoRR, abs/2501.03771.



Bulhak, A. C. (1996).

On the simulation of postmodernism and mental debility using recursive transition networks.

Monash University Department of Computer Science.



Cabanac, G. and Labbé, C. (2021).

Prevalence of nonsensical algorithmically generated papers in the scientific literature.

Journal of the Association for Information Science and Technology, 72(12):1461–1476.



Cabanac, G., Labbé, C., and Magazinov, A. (2021).

Tortured phrases: A dubious writing style emerging in science. evidence of critical issues affecting established journals.



Clause, A., Badalova, F., Cabanac, G., and Mayr, P. (2025).

Unveiling tortured phrases in humanities and social sciences.

CoRR, abs/2502.04944.



Csiszar, A. (2020).

Citation indexes for science.

Science, 122(3159):108–111.



Gingras, Y. (2019).

The Transformation of the Scientific Paper: From Knowledge to Accounting Unit, pages 43–55. MIT press.



Gingras, Y. and Khelifaoui, M. (2021).

L'effet sigaps : la recherche médicale française sous l'emprise de l'évaluation comptable.

Zilsel, 8(1):144–174.



Hirsch, J. E. (2005).

An index to quantify an individual's scientific research output.

Proceedings of the National Academy of Science, 102:16569–16572.



Hvistendahl, M. (2013).

China's publication bazaar.

Science, 342(6162):1035–1039.



Ibrahim, H., Debicki, M., Rahwan, T., and Zaki, Y. (2024).

Big tech dominance despite global mistrust.

IEEE Trans. Comput. Soc. Syst., 11(3):3741–3752.



Kobak, D., González-Márquez, R., Horvát, E., and Lause, J. (2024).

Delving into chatgpt usage in academic writing through excess vocabulary.

CoRR, abs/2406.07016.



Kwon, D. (2024).

AI-generated images threaten science — here's how researchers hope to spot them.

Nature, 635(8038):268–269.



Labbé, C. (2010).

He antike, one of the great stars in the scientific firmament

Semi-automated fact-checking of nucleotide sequence reagents in biomedical research publications: The seek & blastn tool.

PLOS ONE, 14(3):1–22.



Labbé, C. and Labbé, D. (2013).

Duplicate and fake publications in the scientific literature: how many scigen papers in computer science?

Scientometrics, 94(1):379–396.



Liu, Q., Barhomi, A., and Labbé, C. (2024).

Miscitations in scientific papers: dataset and detection.

A workshop paper for The Bibliometric-enhanced Information Retrieval workshop series (BIR 2024).



Lopez-Cozar, E. D., Robinson-García, N., and Torres-Salinas, D. (2012).

Manipulating google scholar citations and google scholar metrics: Simple, easy and tempting.

arXiv preprint arXiv:1212.0638.



Lu, C., Lu, C., Lange, R. T., Foerster, J. N., Clune, J., and Ha, D. (2024).

The AI scientist: Towards fully automated open-ended scientific discovery.

CoRR, abs/2408.06292.



Oviedo-García, M. Á. (2024).

The review mills, not just (self-)plagiarism in review reports, but a step further.

Scientometrics, 129(9):5805–5813.



Ozturk, Y., Bordignon, F., Labbé, C., and Portet, F. (2024).

Assessing the Impact of Retracted Articles on the Reliability of Citing Literature: A Mixed-Based Proposal.

In *MET STI 2024*, On-line, France, Zenodo.



Richardson, R. (2024).

Engineering the world's highest cited cat, larry.

