Exploration of Tortured Phrases Detection

Cambodia Academy of Digital Technology

Institute of Digital Research and Innovation

Digital Research and Development Center

Puthineath Lay

CADT IDRI Institute of Digital Research & Innovation

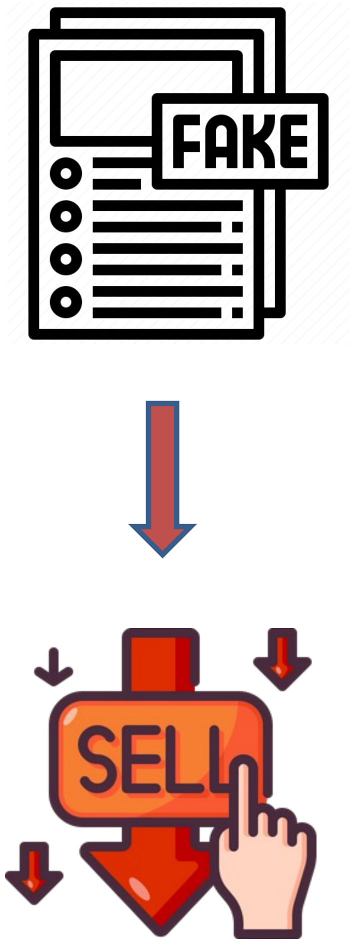
Université Grenoble Alpes

Project Title: Exploration of Tortured Phrases Detection

Introduction:

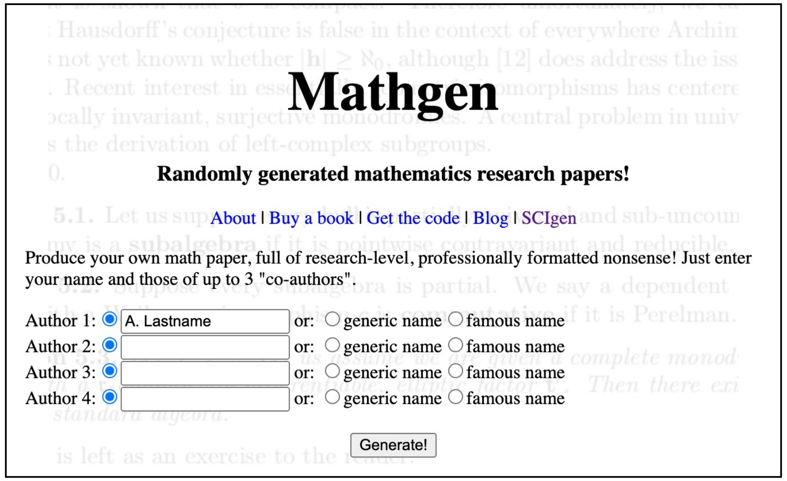
“As a result, meaningless randomly generated scientific papers end up being served and sometimes sold by various publishers with a prevalence estimated to 4.29 papers every one million papers”.

-Cabanac & Labbé (2021)



Project Title: Exploration of Tortured Phrases Detection

Introduction:



Mathgen

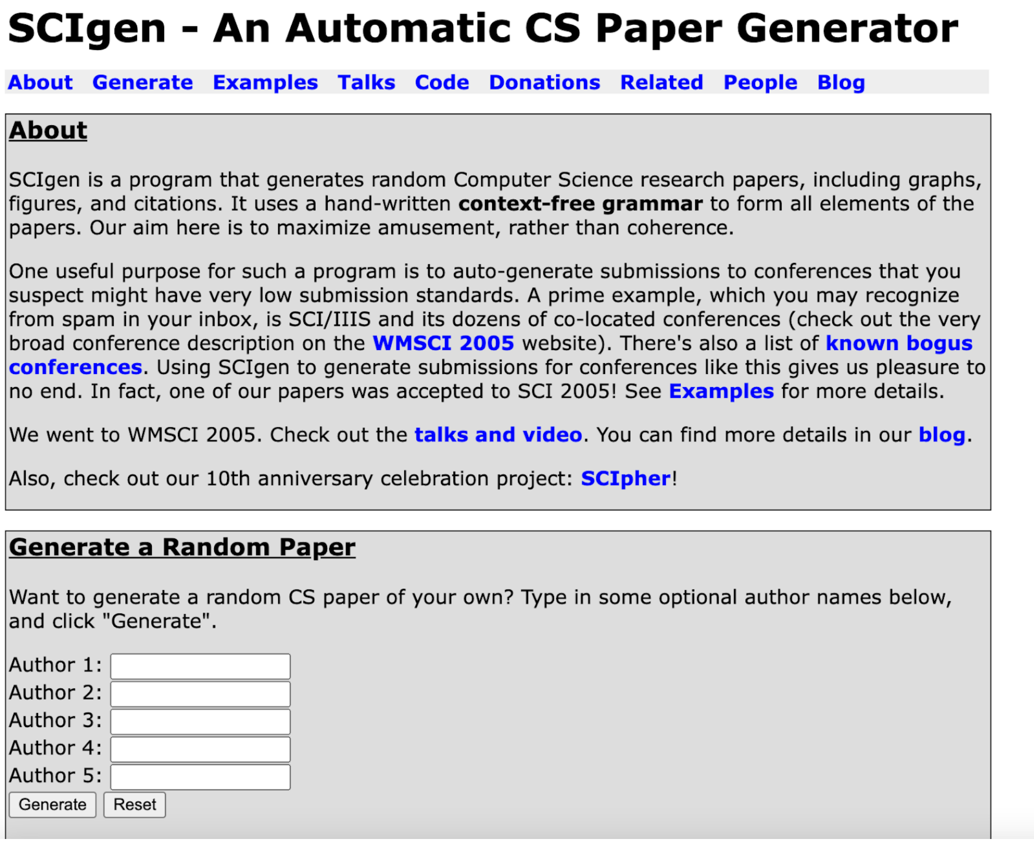
Randomly generated mathematics research papers!

About | Buy a book | Get the code | Blog | SCIgen

Produce your own math paper, full of research-level, professionally formatted nonsense! Just enter your name and those of up to 3 "co-authors".

Author 1: ● A. Lastname or: ○ generic name ○ famous name
Author 2: ● or: ○ generic name ○ famous name
Author 3: ● or: ○ generic name ○ famous name
Author 4: ● or: ○ generic name ○ famous name

Generate!



SCIgen - An Automatic CS Paper Generator

About Generate Examples Talks Code Donations Related People Blog

About

SCIgen is a program that generates random Computer Science research papers, including graphs, figures, and citations. It uses a hand-written **context-free grammar** to form all elements of the papers. Our aim here is to maximize amusement, rather than coherence.

One useful purpose for such a program is to auto-generate submissions to conferences that you suspect might have very low submission standards. A prime example, which you may recognize from spam in your inbox, is SCI/IIIS and its dozens of co-located conferences (check out the very broad conference description on the **WMSCI 2005** website). There's also a list of **known bogus conferences**. Using SCIgen to generate submissions for conferences like this gives us pleasure to no end. In fact, one of our papers was accepted to SCI 2005! See **Examples** for more details.

We went to WMSCI 2005. Check out the **talks and video**. You can find more details in our **blog**.

Also, check out our 10th anniversary celebration project: **SCIpher**!

Generate a Random Paper

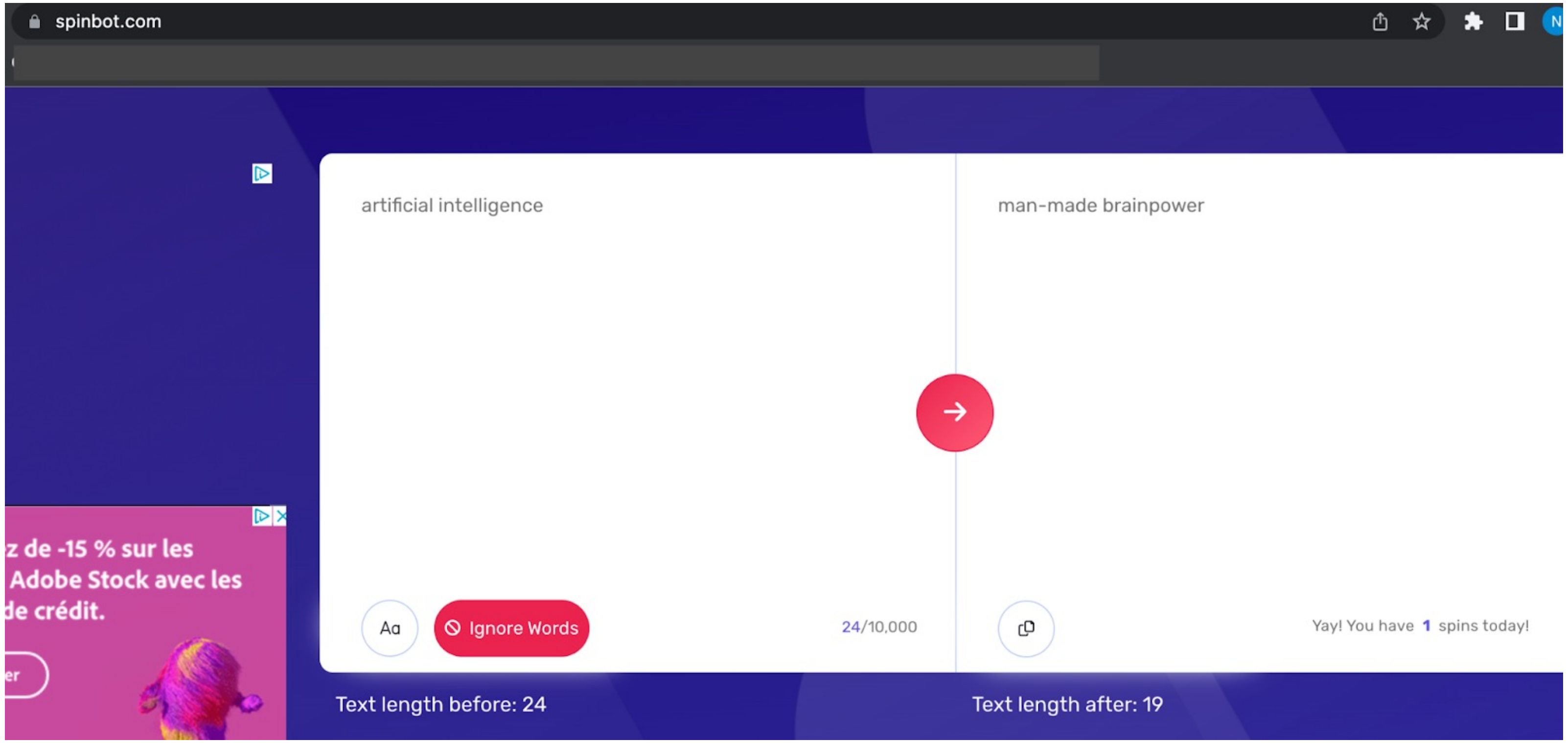
Want to generate a random CS paper of your own? Type in some optional author names below, and click "Generate".

Author 1:
Author 2:
Author 3:
Author 4:
Author 5:

Generate Reset

Generated mathematics research papers websites

Introduction:



spinbot.com

artificial intelligence

man-made brainpower

Text length before: 24

Text length after: 19

Artificial Intelligence -> man-made brainpower

Introduction:

Tortured phrases & **Expected phrases**

Tortured phrases are unexpected weird phrases instead of the established ones, such as “counterfeit consciousness” or “man-made brain power” instead of “artificial intelligence” which is **Expected phrase**.

Example of **Tortured phrases** & **Expected phrases**

- Unused York: New York
- innocent Bayes : naive Bayes
- immature nations : developing countries

Objectives:

My study is to detect the new (unlisted) kind of the tortured phrases in the sentences automatically

Example: It is commonly acknowledged that FDI is one of the essential wellsprings of capital inflow and driving components of financial development in many creating nations.

- creating nations is the tortured phrase.
- developing countries is the expected phrase.

Datasets:

1. Tortured Phrases (Canabac et el., 2021): Human Evaluation on Tortured Phrases

2. Contexts contain *tortured phrases* (Wahle et al., 2021): Machine-paraphrased by Spinbot and SpinnerChief Evaluation on Tortured Phrases

Techniques:

- Machine learning classification:
  - Random Forest classifier
  - Perceptron classifier
  - Transformer-based classifier

- Cosine similarity of words: by using GloVe and BERT to check the cosine score of the phrases.

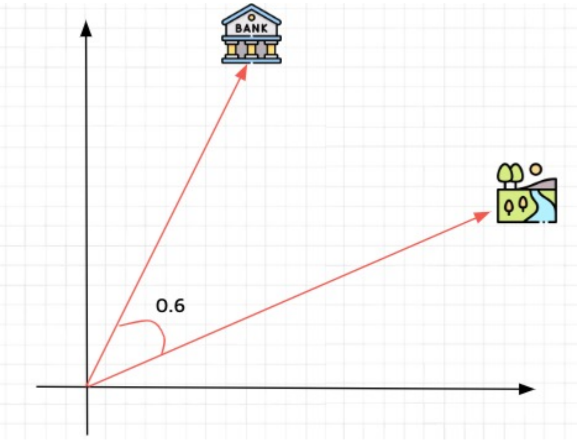
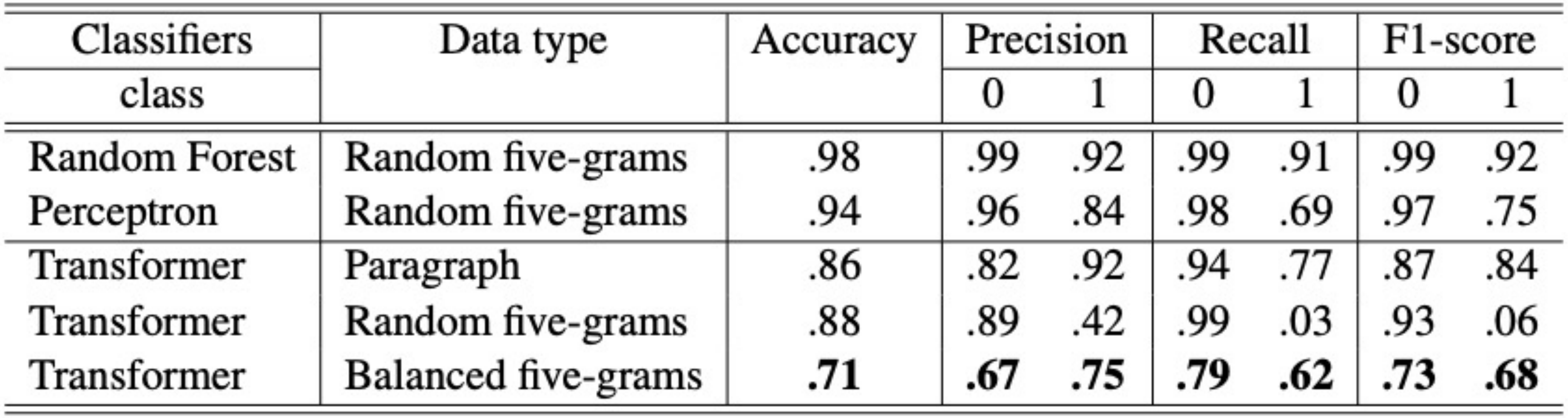


Figure1: example of cosine score using BERT of “bank” from “river bank” and “bank account”.

Results:



Classifiers	Data type	Accuracy	Precision		Recall		F1-score	
class			0	1	0	1	0	1
Random Forest	Random five-grams	.98	.99	.92	.99	.91	.99	.92
Perceptron	Random five-grams	.94	.96	.84	.98	.69	.97	.75
Transformer	Paragraph	.86	.82	.92	.94	.77	.87	.84
Transformer	Random five-grams	.88	.89	.42	.99	.03	.93	.06
Transformer	Balanced five-grams	**.71**	**.67**	**.75**	**.79**	**.62**	**.73**	**.68**

Table 1: Classfication results

Project Title: Exploration of Tortured Phrases Detection

Results:

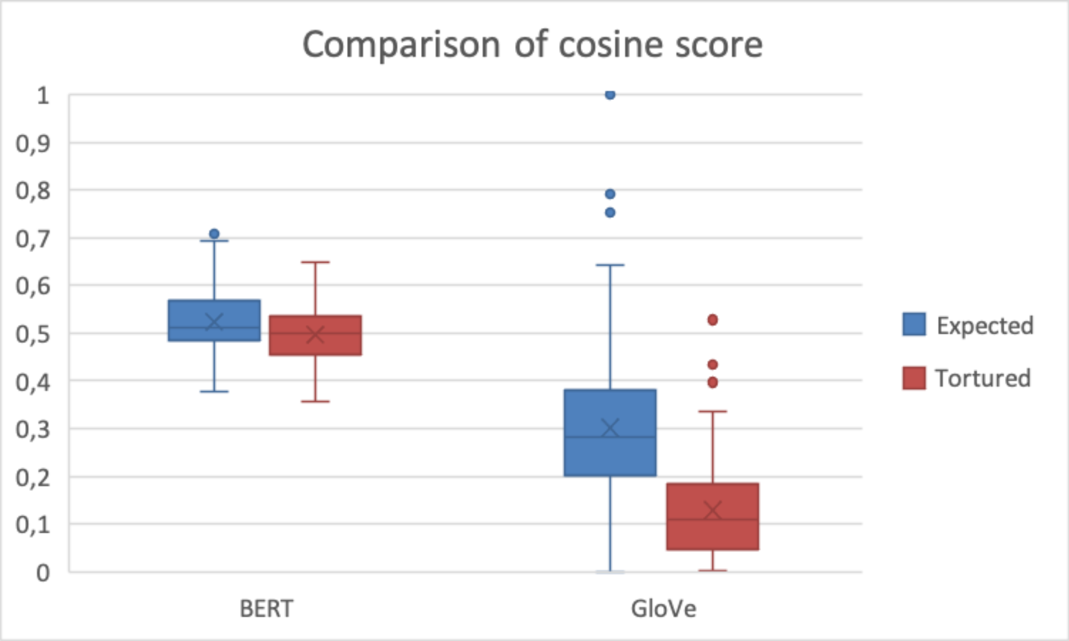


Figure 1: Result of cosine score comparison using GloVe and BERT

CADT
IDRI Institute of Digital Research & Innovation

Outcomes:

It can help users/researchers avoid fraudulent papers. It can be adapt to other languages such as Khmer, Thai, and so on.

ភាសាខ្មែរ

ภาษาไทย

ພາສາລາວ

Project Title: Exploration of Tortured Phrases Detection

Impacts:

- A model can detect machine-generated text.
- The source code and the data is available online.

Conclusion:

- Transformer-based classifier and cosine score using pre-trained embedding perform noticeable results.
- Need more dataset for the training
- Need to experiment on other classifiers and word embedding such as word2vec, fasttext.

References:

- Cabanac, G., & Labbé, C. (in press). Prevalence of nonsensical algorithmically generated papers in the scientific literature. Journal of the Association for Information Science and Technology. doi: 10.1002/asi.24495

- Guillaume Cabanac, Cyril Labbé, and Alexander Magazinov. Tortured phrases: A dubious writing style emerging in science. evidence of critical issues affecting established journals. CoRR, abs/2107.06751, 2021.

- Jan Wahle, Terry Ruas, Tomas Foltynek, Norman Meuschke, and Bela Gipp. Identifyingmachine-paraphrased plagiarism. 03 2021

- Puthineath Lay, Martin Lentschat, and Cyril Labbe. 2022. Investigating the detection of Tortured Phrases in Scientific Literature. In Proceedings of the Third Workshop on Scholarly Document Processing, pages 32–36, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Thank you for your attention!

Cambodia Academy of Digital Technology

Institute of Digital Research and Innovation

Digital Research and Development Center

Puthineath Lay
Email: puthineath.lay@cadt.edu.kh
Phone Number: +855 (0)69 83 53 63

CADT IDRI Institute of Digital Research & Innovation

More details

Cosine score(tortured phrases and expected phrases matching by ID):

- All GloVe: https://htmlpreview.github.io/?https://github.com/Puthineath/Detection/blob/main/cosine2tokens_glove.html
- Glove:
  https://htmlpreview.github.io/?https://github.com/Puthineath/Detection/blob/main/cosine2tokens_glove_final.html
- BERT : https://htmlpreview.github.io/?https://github.com/Puthineath/Detection/blob/main/cosine2tokens_bert_final.html