

Ученые обнаружили математические "отпечатки пальцев" писателей

Шведские ученые разработали новый способ выявлять автора текста - они установили, что роль "отпечатков пальцев" писателей может играть частота встречаемости новых слов в их текстах.

В статье, опубликованной в четверг в научном издании New Journal of Physics, группа шведских физиков из университета Умео под руководством Себастьяна Бернгардсона (Sebastian Bernhardsson) описала новый метод, который позволяет на основе статистических данных определить автора текста.

Исследователи проверяли, как в текстах трех писателей - Томаса Харди, Генри Мелвилла и Дэвида Лоуренса - реализуется так называемый закон Ципфа. Этот закон, открытый в 1935 году лингвистом Джорджем Ципфом (George Kingsley Zipf), гласит, что частота какого-либо слова в текстах обратно пропорциональна его рангу - месту в списке слов текста, отсортированных по частоте.

Так, например, второе по частоте слово будет встречаться в тексте примерно в два раза реже, чем первое, третье - в три раза реже и так далее.

Шведские физики в своей статье показали, что этот закон не так универсален, как считал Ципф. Они обнаружили, что частота появления новых слов по мере роста объема текста меняется у разных авторов по-разному, причем эта закономерность не зависит от конкретного текста, а только от автора.

Статистический анализ показал, что закономерность остается постоянной в любых текстах одного и того же автора - романах, главах из романов, рассказах, и может служить своеобразными "отпечатками пальцев".

Авторы исследования, наблюдая за этими статистическими закономерностями, выдвинули идею так называемой "метакниги" - воображаемого бесконечного текста, в котором описан мир глазами того или иного автора. "Создавая произведение, автор "вытаскивает" куски текста из этой большой "материнской книги" и перекладывает их на бумагу, сохраняя, однако, частотные характеристики концептов в этой метакниге", - пишут ученые.

Статистические методы определения авторства известны довольно давно, и шведские исследователи просто предложили еще один, относительно простой способ, сказала РИА Новости лингвист Елизавета Былина из Утрехтского университета (Нидерланды).

"Традиционно считается, что авторский стиль хорошо характеризуется распределением служебных слов, и другими незначимыми, и потому плохо заметными глазу, а хорошо заметными статистике характеристиками - средняя длина предложения, количество вводных слов", - сказала собеседница агентства.

По ее словам, Бернгардсон и его соавторы попытались представить некоторую частотную "карту" произведений, и нашли, что она постоянна для каждого автора, даже без всяких других "хитростей".

"Это хороший результат, но не сенсационный", - считает лингвист.

Также читайте эксклюзивную информацию в соцсетях:

- [Телеграм](#)

- [ВКонтакте](#)

Связаться с редакцией Бабра:

newsbabr@gmail.com

НАПИСАТЬ ГЛАВРЕДУ:

Телеграм: @babr24_link_bot
Эл.почта: newsbabr@gmail.com

ЗАКАЗАТЬ РАССЛЕДОВАНИЕ:

эл.почта: bratska.net.net@gmail.com

КОНТАКТЫ

Бурятия и Монголия: Станислав Цырь
Телеграм: @bur24_link_bot
эл.почта: bur.babr@gmail.com

Иркутск: Анастасия Суворова
Телеграм: @irk24_link_bot
эл.почта: irkbabr24@gmail.com

Красноярск: Ирина Манская
Телеграм: @kras24_link_bot
эл.почта: krasyar.babr@gmail.com

Новосибирск: Алина Обская
Телеграм: @nsk24_link_bot
эл.почта: nsk.babr@gmail.com

Томск: Николай Ушайкин
Телеграм: @tomsk24_link_bot
эл.почта: tomsk.babr@gmail.com

[Прислать свою новость](#)

ЗАКАЗАТЬ РАЗМЕЩЕНИЕ:

Рекламная группа "Экватор"
Телеграм: @babrobot_bot
эл.почта: equatoria@gmail.com

СТРАТЕГИЧЕСКОЕ СОТРУДНИЧЕСТВО:

эл.почта: babrmarket@gmail.com

[Подробнее о размещении](#)

[Отказ от ответственности](#)

[Правила перепечаток](#)

[Соглашение о франчайзинге](#)

[Что такое Бабр24](#)

[Вакансии](#)

[Статистика сайта](#)

[Архив](#)

[Календарь](#)

[Зеркала сайта](#)