



empenoso

3 часа назад

Как я пытался ускорить анализ 12 000 комментариев с помощью GPU за 50 тысяч, но победил процессор

4 мин 710

Open source*, Настройка Linux*, Python*, Контент и копирайтинг*

История о том, как я хотел провести анализ комментариев, а в итоге получил неожиданный, но полезный опыт с локальным AI.

	A	B	C	D	E	F	G
1	Ресурс	Кол-во	Начало	Конец	Тираж (печатн.)	Просмотры (веб)	Комментарии (веб)
2	https://habr.com	113	2019-11	2025-09	0,0	2 192 876,0	4 497,0
3	https://pikabu.ru	61	2024-09	2025-09	0,0	506 243,0	726,0
4	https://smart-lab.ru	50	2024-09	2025-09	0,0	453 955,0	1 998,0
5	https://www.youtube.com	22	2013-02	2025-09	0,0	182 381,0	22,0
6	https://alenka.capital	1	2025-09	2025-09	0,0		0,0
7	https://fkviking.timepad.ru	1	2025-09	2025-09	0,0		0,0
8	https://rationalanswer.club	7	2024-11	2025-08	0,0	45 078,0	39,0
9	https://vc.ru	11	2019-11	2025-08	0,0	11 447,0	9,0
10	https://t.me	12	2019-03	2024-12	0,0	79 200,0	122,0
11	https://www.1tv.ru	1	2024-12	2024-12	0,0		0,0
12	https://sprut.ai	3	2024-05	2024-11	0,0		0,0
13	https://t-j.ru	49	2020-04	2024-10	0,0	2 300 823,0	4 186,0
14	https://github.com	10	2019-09	2023-06	0,0	0,0	0,0
15	https://podcast.ru	1	2022-11	2022-11	0,0		0,0
16	https://medium.com	1	2020-02	2020-02	0,0		0,0
17	https://zen.yandex.ru	1	2019-12	2019-12	0,0		0,0
18	https://3dtoday.ru	13	2018-09	2018-11	0,0	432 787,0	553,0
19	Журнал Инновации и инвестиции	2	2018-04	2018-05	600,0		0,0
20	Сборник ПНИПУ	7	2012-04	2017-05	700,0		0,0
21	https://z-wave.ru	16	2014-11	2015-04	0,0		0,0

LynxReport: учёт публикаций [Node.js] ✓

Недавно передо мной встала задача собрать все положительные комментарии к моим статьям. Веду их учёт в таблице, и там уже вполне серьёзные цифры — больше 300 строк и свыше 10 тысяч комментариев. Основные площадки, где я публикуюсь, выглядят так:

- Хабр — 4 497 комментариев
- Т-Ж — 4 186
- Смартлаб — 1 998
- Пикабу — 726

Вручную искать в этом массиве текста слова поддержки — долго и нудно, а главное — совершенно не масштабируется. Так родилась идея: поручить всё локальной нейросети для анализа тональности. Заодно я хотел на практике разобраться с моделями на основе BERT.

Для этой цели у меня был, как мне казалось, идеальный инструмент: компьютер с Ubuntu 24 LTS, 16 ГБ

РЕКЛАМА



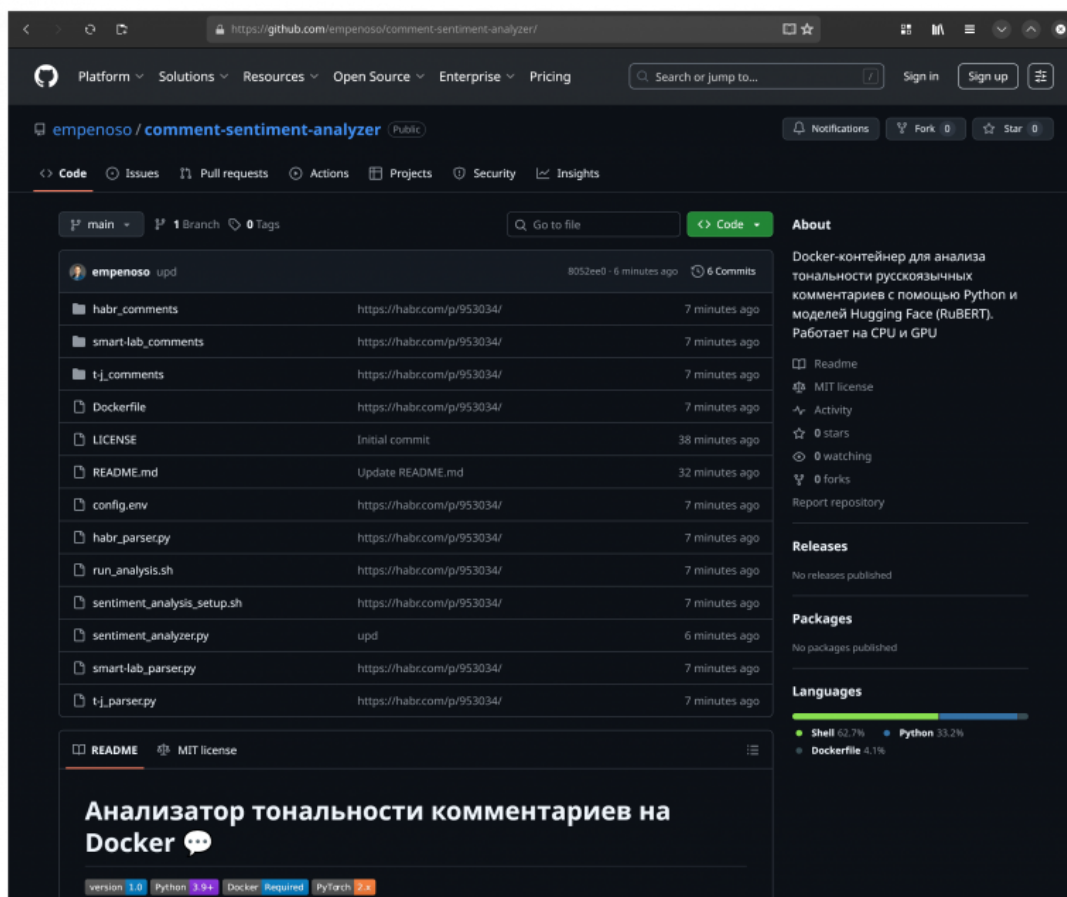
Ждем на HighLoad++

10 треков, 150+ спикеров

анализ пройдёт молниеносно. Но на деле именно GPU стал главной точкой отказа — пришлось всё считать на CPU.

Код на [GitHub](#).

Что делает мой скрипт на GitHub



<https://github.com/empenoso/comment-sentiment-analyzer/>

Прежде чем бросаться в бой с CUDA, нужно было подготовить данные. Комментарии разбросаны по разным сайтам без возможности экспорта, поэтому я написал несколько парсеров, которые собирают тексты в единый JSON-формат — один файл на статью.



Ждем на HighLoad++
10 треков, 150+ спикеров

JSON Хабр

Парсеры я оставил приватными, так как они заточены под мои задачи, но [ядро системы выложил на GitHub](#).

Скрипт `sentiment_analyzer.py` берёт JSON-файлы, подготовленные на первом этапе, и пропускает текст каждого комментария через предварительно обученную нейросетевую модель [cointegrated/rubert-tiny-sentiment-balanced](#). Модель определяет эмоциональную окраску текста — позитивную, негативную или нейтральную.

Скрипт задуман как универсальный инструмент, поэтому основные параметры вынесены в `config.env`.

- `MODEL_NAME`: можно указать любую другую модель с Hugging Face.
- `DEVICE`: позволяет выбрать, на чём производить вычисления — на `cpu` или `cuda` (GPU).
- `POSITIVE_THRESHOLD`: порог уверенности модели, чтобы отнести комментарий к позитивным.
- `EXCLUDE_AUTHORS`: список авторов, чьи комментарии нужно игнорировать (например, мои собственные ответы).
- `MIN_COMMENT_LENGTH`: отсеивает слишком короткие и неинформативные комментарии.

Оркестрация с помощью Docker и Shell



Ждем на HighLoad++
10 треков, 150+ спикеров

1. `sentiment_analysis_setup.sh`: этот скрипт — для первоначальной настройки. Он проверяет систему, устанавливает Docker и NVIDIA Container Toolkit, создаёт необходимые папки и конфигурационный файл. Запустив его один раз, вы подготавливаете окружение для дальнейшей работы.
2. `run_analysis.sh`: простой скрипт для запуска анализа. Он читает конфигурацию из `config.env` и запускает Docker-контейнер с нужными параметрами.

На практике это сводится к трём шагам: подготовка системы через `setup`-скрипт, сбор комментариев парсерами и запуск анализа через `run_analysis.sh`.

Все найденные позитивные комментарии скрипт аккуратно складывает в текстовые файлы.

Успех на CPU и урок о масштабировании

После череды падений с CUDA и финальной ошибки «No kernel image» пришлось смириться: GPU в проекте не будет. Я открыл `config.env`, поменял `DEVICE=cuda` на `DEVICE=cpu` и нажал `save`.



Ждем на HighLoad++
10 треков, 150+ спикеров

RuBERT-tiny справился за полторы минуты — все 10 000 комментариев были разобраны. Вся похвала и поддержка оказались в аккуратных текстовых файлах.

На финишной прямой я переписал логику сохранения. Вместо сотен мелких JSON теперь формируется один аккуратный текстовый файл для каждой площадки. Структура вывода проста:

Автор: Андрей Мищенко

Дата: 2025-06-06T11:24:20.551316+03:00

Текст: Полезные формулы, спасибо!

Ссылка: <https://t-j.ru/guide/excel-kotirovki/#c2857460>

Автор: Whalerman

Дата: 2025-09-09T07:40:00.000Z

Текст: Михаил, спасибо! Хорошие и полезные посты!

Ссылка: <https://smart-lab.ru/blog/1202442.php#comment18571129>

Автор: DashBerlin

Дата: 2025-08-23T00:18:43.000Z

Текст: Впервые решил заглянуть в подобный обзор, порадовала позиция количество закладок, интересно, после эти публикации читаются. Я подписан на автора, он пачками сохраняет статьи в закладки, какой процент он потом перечитывает из этого. Спасибо за обзор))

Ссылка: https://habr.com/ru/articles/933806/comments/#comment_28742672



Ждем на HighLoad++
10 треков, 150+ спикеров

Я ожидал, что GPU будет обрабатывать тысячи комментариев в секунду, но реальность оказалась прагматичнее — CPU выдал скорость около 110 комментариев в секунду.

Этот опыт наглядно показал: локальный AI на процессоре — отличное решение для задач исследователя-одиночки. Но если бы у меня был миллиард строк, этот подход бы провалился. Это инструмент для прототипирования и персональных проектов, а не для Big Data.

Заключение: главный урок для AI-энтузиастов

Эта история с видеокартой преподала мне два важных урока.

Во-первых, самое новое оборудование — не всегда самое лучшее. Моя RTX 5060 Ti с передовой архитектурой Blackwell оказалась настолько свежей, что стабильный PyTorch просто не умел с ней работать. В погоне за технологиями легко обогнать экосистему и остаться с мощным, но бесполезным инструментом (я знаю [про обходной путь](#)). Иногда проверенная карта предыдущего поколения — более разумный выбор.

Во-вторых, Docker — это не просто среда для запуска, а настоящая страховка. Он позволил мне безболезненно переключаться между конфигурациями и быстро откатиться на CPU, когда GPU подвёл. Именно изоляция в контейнере спасла проект, позволив проиграть «битву за CUDA», но всё равно выиграть войну.

В итоге, мой проект заработал не благодаря дорогой видеокарте, а вопреки ей. Процессор решил задачу быстрее, чем я успел допить чай. Это доказывает, что локальный AI — не удел облачных гигантов. Он вполне доступен на обычных ПК, если подходить к делу прагматично и помнить, что иногда самое простое решение — самое верное.

Автор: Михаил Шардин

 [Моя онлайн-визитка](#)

 [Telegram «Умный Дом Инвестора»](#)

7 октября 2025

Теги: [cuda](#), [sm_120](#), [blackwell](#), [rtx 5060](#), [bert](#)

Хабы: [Open source](#), [Настройка Linux](#), [Python](#), [Контент и копирайтинг](#)

Редакторский дайджест

Присылаем лучшие статьи раз в месяц



Оставляя свою почту, я принимаю [Политику конфиденциальности](#) и даю согласие на получение рассылок



218

Карма

64.4

Общий рейтинг

Михаил Шардин [@empenoso](#)



Ждем на HighLoad++
10 треков, 150+ спикеров

[Сайт](#) [Сайт](#) [GitHub](#)[Комментарии 6](#)

Публикации

[ЛУЧШИЕ ЗА СУТКИ](#)[ПОХОЖИЕ](#)**DRoman0v**

20 часов назад

Воскресная барахолка под Валенсией: винтажная аудиотехника, ноутбуки и много чего ещё

7 мин

3.2K

+45

4

2

**interpres**

14 часов назад

Отвлекать разработчиков ПО намного вреднее, чем считает большинство менеджеров

Простой

7 мин

6.3K



Ждем на HighLoad++
10 треков, 150+ спикеров

 +40 43 13

alex0x08

20 часов назад

WiFi, который не ловил

 Простой 4 мин 5.8K

Тutorial

 +37 41 26

IgorAnokhin

17 часов назад

Асинхронность vs. многопоточность: что выживет в эпоху No GIL?

 14 мин 4.8K

Мнение

 +34 56 27

RationalAnswer

22 часа назад

Шатдаун правительства США, а также Грокипедия от Илона Маска

 8 мин 6.6K

Дайджест

 +31 10 7

Erwinmal

18 часов назад

Pasta e basta! Как паста и макароны стали национальным блюдом Италии и покорили Россию и США? Часть 3

 Простой 14 мин 1.7K

Ретроспектива

 +27 14 0

dalerank

10 часов назад

Сам себе breakpoint

 Простой 15 мин 2.1K +25 26 9

adushein

18 часов назад

Пятьдесят оттенков отказа: стоп слово — Zonal Shift

 10 мин 2.1K +25 26 9

Ждем на HighLoad++
10 треков, 150+ спикеров

 0



nikgerasimenko

19 часов назад

Kandinsky Image научился генерировать изображения с надписями на русском

Простой

3 мин

2.2K

Обзор

+21

8

2



PatientZero

15 часов назад

У меня нет рта, но я должен выводить эмодзи морского конька

Средний

9 мин

2.1K

Обзор

Перевод

+19

7

15

Гибкое управление секретами, прозрачность и бесшовная интеграция с DevOps: как надёжно защитить данные

Турбо

Показать еще

ИСТОРИИ



10 признаков, что вы работаете не там



На мороз!

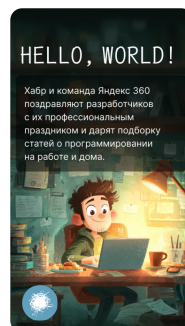
История айтишников из самых холодных уголков планеты



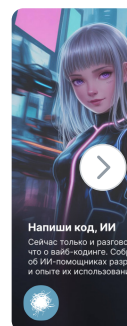
Прогулка по IT Elements



Годнота из блогов компаний



С Днём программиста!



Made in AI

ВАКАНСИИ

Deep Learning Engineer (GigaChat Prod)

от 350 000 Р · Сбер · Москва

Linux администратор HPC стека

от 200 000 до 400 000 Р · Сбер · Москва



Ждем на HighLoad++

10 треков, 150+ спикеров

Linux администратор HPC стека

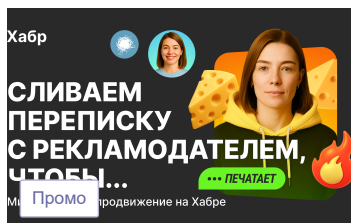
от 200 000 до 350 000 Р · Сбер · Москва

Python разработчик

от 280 000 до 380 000 Р · Selecty · Можно удаленно

[Больше вакансий на Хабр Карьере](#)

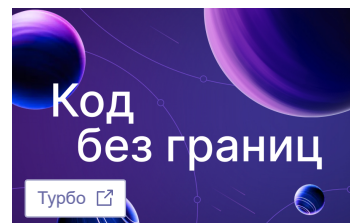
МИНУТОЧКУ ВНИМАНИЯ



Как продвигаться на Хабре, если вы — не IT-бренд

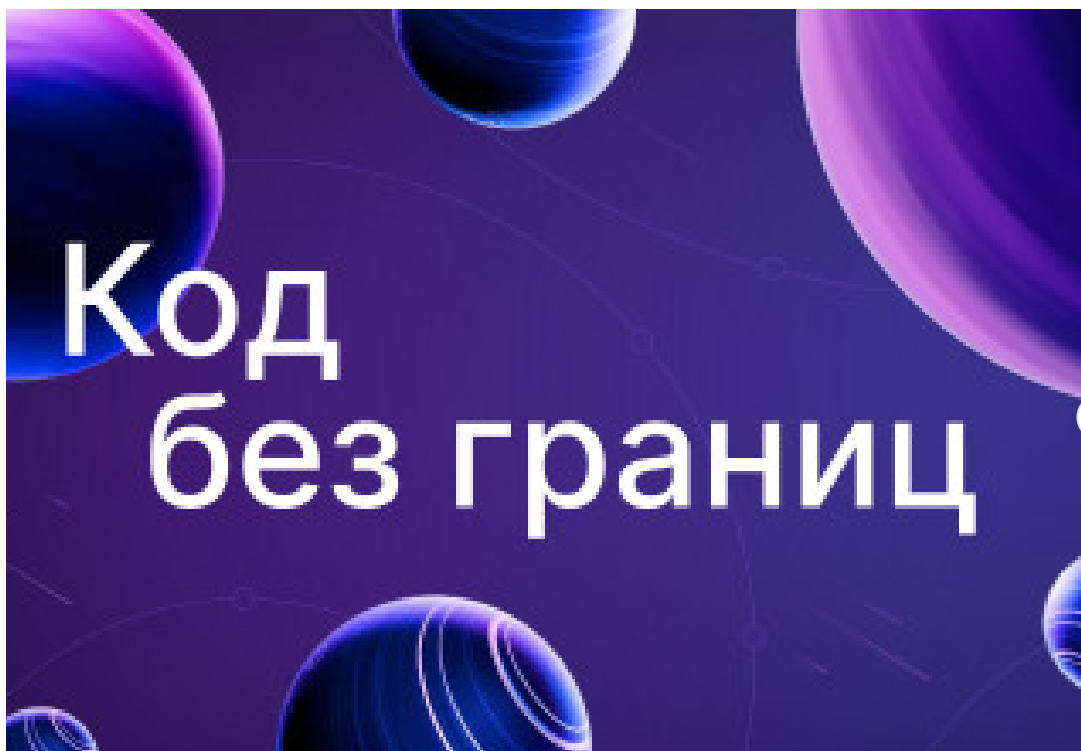


Каков твой выбор?



Отправь свой open source проект на конкурс и выиграй грант

БЛИЖАЙШИЕ СОБЫТИЯ



3 сентября – 31 октября

Программа грантов для развития open source проектов «Код без границ»

Онлайн

Разработка

[Больше событий в календаре](#)



Ждем на HighLoad++

10 треков, 150+ спикеров

Хабр



Настройка языка

Техническая поддержка

© 2006–2025, Habr



Ждем на HighLoad++
10 треков, 150+ спикеров