

Во время посещения сайта вы соглашаетесь с использованием файлов [cookie](#)

Хорошо



Михаил Шардин ★

личный блог



07 октября 2025, 04:29

+ Подписаться

Как я строил “AI для трейдинга”, а нашёл способ собирать похвалу

Меня иногда упрекают, что я пишу не про трейдинг, а про какие-то «технические игрушки». Но на самом деле все эти проекты — из одной экосистемы. Ведь алгоритмический трейдинг начинается не с покупки кнопки «BUY», а с умений собирать, очищать и анализировать данные.

```
mike@linux-pc: ~/SynologyProjects/2025_10_Тональность комментариев

=====
== CUDA ==
=====

CUDA Version 12.4.1

Container image Copyright (c) 2016-2023, NVIDIA CORPORATION & AFFILIATES. All rights reserved.

This container image and its contents are governed by the NVIDIA Deep Learning Container License.
By pulling and using the container, you accept the terms and conditions of this license:
https://developer.nvidia.com/ngc/nvidia-deep-learning-container-license

A copy of this license is made available in this container at /NGC-DL-CONTAINER-LICENSE for your convenience.

WARNING: The NVIDIA Driver was not detected. GPU functionality will not be available.
Use the NVIDIA Container Toolkit to start this container with GPU support; see
https://docs.nvidia.com/datacenter/cloud-native/ .

=====
🗨️ АНАЛИЗАТОР ТОНАЛЬНОСТИ КОММЕНТАРИЕВ
=====

⚙️ КОНФИГУРАЦИЯ
=====
📁 Модель: coIntegrated/rubert-tiny-sentiment-balanced
🖨️ Устройство: cpu
📊 Порог позитивности: 0.8
👤 Исключить авторов: ['empenoso', 'Михаил Шардин']
📏 Мин. длина: 20 символов
📁 Входные папки: ['habr_comments', 'smart-lab_comments', 't-j_comments']
=====

🔍 В 'habr_comments' найдено 106 json-файлов.
🔍 В 'smart-lab_comments' найдено 49 json-файлов.
🔍 В 't-j_comments' найдено 48 json-файлов.

📁 Всего для обработки: 203 файлов.
📦 Загрузка модели coIntegrated/rubert-tiny-sentiment-balanced...
✅ Модель загружена на CPU

=====
🚀 НАЧАЛО АНАЛИЗА
=====
```

Откройте счёт в ВТБ Мои Инвестиции

Введите текст комментария

с финансовыми новостями: собираем данные, структурируем, прогоняем через модель, выделяем позитив и негатив. В трейдинге это может стать элементом новостного сканера или инструмента для оценки тональности рынка. А в моём случае — это просто удобный способ собрать добрые слова, за которые я всем благодарен.

	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://alenska.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

Откройте счёт в ВТБ Мои Инвестиции

BERT.

Для этой цели у меня был, как мне казалось, идеальный инструмент: компьютер с Ubuntu 24 LTS, Docker и мощной NVIDIA GeForce RTX 5060 Ti на 16 ГБ. Казалось, что с RTX 5060 Ti на 16 ГБ 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

The screenshot shows the GitHub repository page for 'empenoso / comment-sentiment-analyzer'. The repository is public and has 0 forks and 0 stars. The main branch is 'main'. The repository contains several files and folders, including 'habr_comments', 'smart-lab_comments', 'tj_comments', 'Dockerfile', 'LICENSE', 'README.md', 'config.env', 'habr_parser.py', 'run_analysis.sh', 'sentiment_analysis_setup.sh', 'sentiment_analyzer.py', 'smart-lab_parser.py', and 'tj_parser.py'. The repository is described as a Docker container for analyzing the sentiment of Russian comments using Python and Hugging Face (RuBERT) models. It works on both CPU and GPU. The repository is licensed under MIT. The repository is also available as a Docker image on Docker Hub.

File	Commit	Time
habr_comments	https://habr.com/p/953034/	7 minutes ago
smart-lab_comments	https://habr.com/p/953034/	7 minutes ago
tj_comments	https://habr.com/p/953034/	7 minutes ago
Dockerfile	https://habr.com/p/953034/	7 minutes ago
LICENSE	Initial commit	38 minutes ago
README.md	Update README.md	32 minutes ago
config.env	https://habr.com/p/953034/	7 minutes ago
habr_parser.py	https://habr.com/p/953034/	7 minutes ago
run_analysis.sh	https://habr.com/p/953034/	7 minutes ago
sentiment_analysis_setup.sh	https://habr.com/p/953034/	7 minutes ago
sentiment_analyzer.py	upd	6 minutes ago
smart-lab_parser.py	https://habr.com/p/953034/	7 minutes ago
tj_parser.py	https://habr.com/p/953034/	7 minutes ago

Анализатор тональности комментариев на Docker

version 1.0 Python 3.9+ Docker Required PyTorch 2.x

About

Docker-контейнер для анализа тональности русскоязычных комментариев с помощью Python и моделей Hugging Face (RuBERT). Работает на CPU и GPU

Releases

No releases published

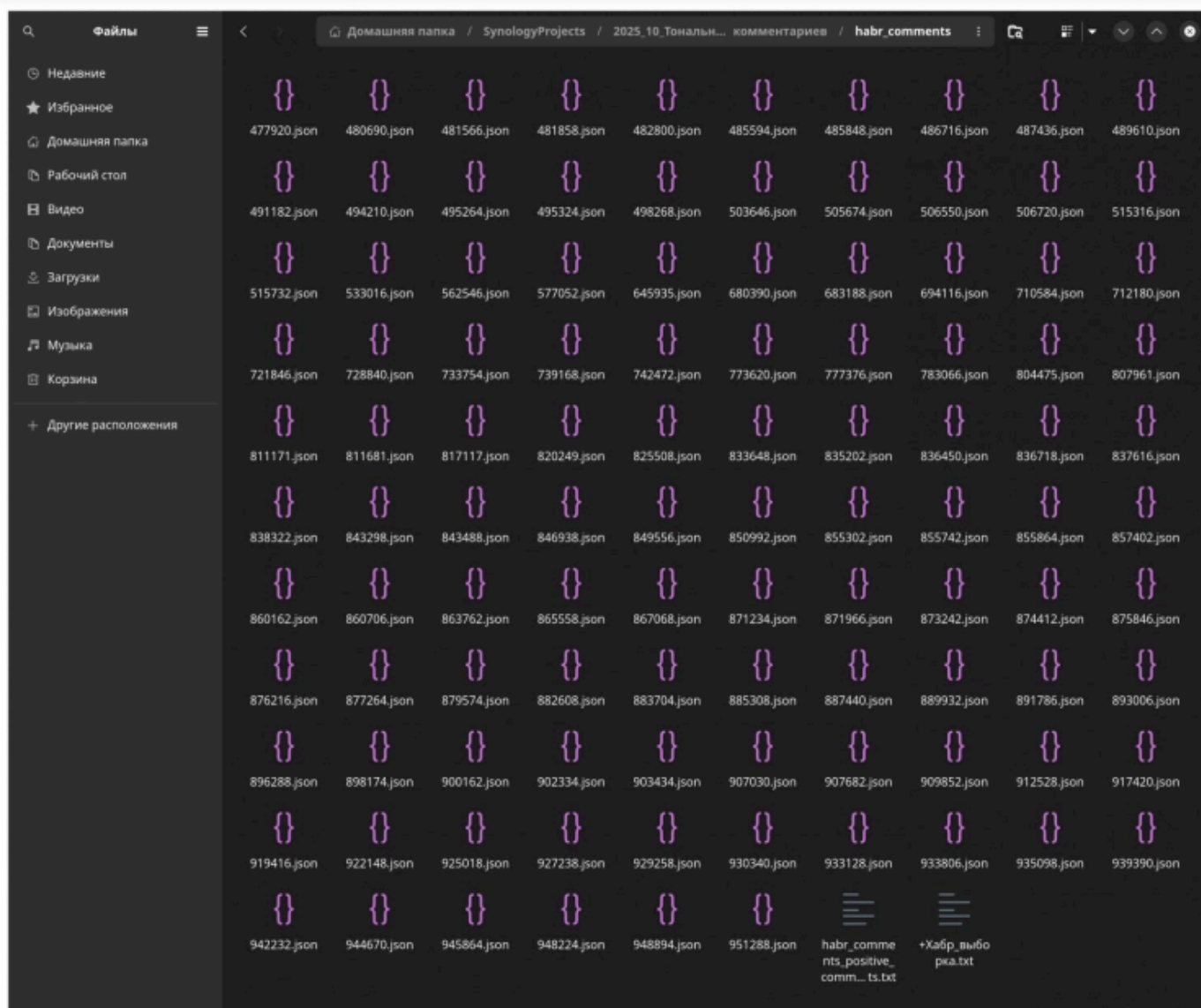
Packages

No packages published

Languages

Shell 62.7% Python 33.2% Dockerfile 4.1%

Откройте счёт в ВТБ Мои Инвестиции



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.

Откройте счёт в ВТБ Мои Инвестиции

- **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

Для удобства запуска и настройки я написал два shell-скрипта:

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.
1. **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`.

Откройте счёт в ВТБ Мои Инвестиции

```
1  #!/usr/bin/env bash
2
3  set -euo pipefail
4
5  # Загружаем конфигурацию
6  if [ -f config.env ]; then
7      source config.env
8  else
9      echo "❌ Файл config.env не найден!"
10     exit 1
11 fi
12
13 echo "🚀 Запуск анализатора тональности в Docker-контейнере..."
14
15 # Проверяем, какое устройство используется
16 if [ "${DEVICE:-cpu}" = "cuda" ]; then
17     echo "Используется GPU для ускорения вычислений."
18     docker run \
19         --rm \
20         --gpus all \
21         -v "$(pwd)":/app \
22         -v "$HOME/.cache/huggingface":"/root/.cache/huggingface" \
23         --env-file config.env \
24         habr-sentiment-analyzer:latest
25 else
26     echo "Используется CPU для вычислений."
27     docker run \
28         --rm \
29         -v "$(pwd)":/app \
30         -v "$HOME/.cache/huggingface":"/root/.cache/huggingface" \
31         --env-file config.env \
32         habr-sentiment-analyzer:latest
33 fi
```

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.

Откройте счёт в ВТБ Мои Инвестиции


```
=====
АНАЛИЗАТОР ТОНАЛЬНОСТИ КОММЕНТАРИЕВ
=====

=====
* КОНФИГУРАЦИЯ
=====

Модель: cointegrated/rubert-tiny-sentiment-balanced
Устройство: cpu
Порог позитивности: 0.8
Исключить авторов: ['empenoso', 'Михаил Шардин']
Мин. длина: 20 символов
Входные папки: ['habr_comments', 'smart-lab_comments', 't-j_comments']
=====

В 'habr_comments' найдено 106 json-файлов.
В 'smart-lab_comments' найдено 49 json-файлов.
В 't-j_comments' найдено 48 json-файлов.

Всего для обработки: 203 файлов.
Загрузка модели cointegrated/rubert-tiny-sentiment-balanced...
✓ Модель загружена на CPU

=====
🚀 НАЧАЛО АНАЛИЗА
=====

[203/203] Обработка 9ce799a4-6a8e-49ee-bd66-c948ba164031.json...

=====
💾 СОХРАНЕНИЕ РЕЗУЛЬТАТОВ
=====

Сохраняю 373 комментария в файл: habr_comments/habr_comments_positive_comments.txt
Сохраняю 182 комментария в файл: smart-lab_comments/smart-lab_comments_positive_comments.txt
Сохраняю 250 комментариев в файл: t-j_comments/t-j_comments_positive_comments.txt

=====
📊 СТАТИСТИКА
=====

Обработано файлов: 203
Найдено позитивных: 805
✓ Анализ завершен! Результаты сохранены в исходных директориях.
=====

mike@linux-pc:~/SynologyProjects/2025_10_Тональность комментариев$
```

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: t-j.ru/guide/excel-kotirovki/#c2857460
```

Откройте счёт в ВТБ Мои Инвестиции

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: habr.com/ru/articles/933806/comments/#comment_28742672 "

Каждый комментарий — отдельный блок, удобно читать и анализировать.

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.

Откройте счёт в ВТБ Мои Инвестиции

Кстати, в этом месяце я снова прилетаю в Москву — 25 октября на конференцию Smart-Lab Conf 2025. Мое выступление состоится в зале № 7 («Спекуляции») в 12:00. Буду рад видеть всех, кто интересуется автоматизацией и новыми подходами к работе с финансовыми данными. До встречи!

AI

торговые роботы

9.5K  9 48 32**Михаил Шардин** Пермь 362  3 511 с 23 января 2019 +HreHDn1F5CZjN...[+ Подписаться](#)

48 КОММЕНТАРИЕВ

Сначала старые **ВалиБакS**

07 октября 2025, 06:45

А как у вас с торговлей то? Успех то тоже такой же как и с программированием? Сколько на текущий момент годовых тогда интересно выбиваете в этом году? Это больше волнует общество трейдеров

[Показать 5 ответов](#)  +6 **Make_hard**

07 октября 2025, 07:32

Привет. Не знаю, было ли целью запустить именно локально, но есть такая штука как google colab. По сути это бесплатный сервис, который выделяет тебе ГПУ и можешь пробовать разные модельки там. Там уже предустановлен пайторч для ГПУ, так что должно работать из коробки) А так — молодец)

[Откройте счёт в ВТБ Мои Инвестиции](#)

вычленить содержательные отзывы, те, которые пригодятся для Вашей торговли.

— Показать 2 ответа



Вазелин

07 октября 2025, 08:28



„AI — не удел облачных гигантов “

Вообще открою тебе тайну. В сфере it можно потратить кучу времени, написать 100500 програмулек, придумать мудреных алгоритмов, запустить аи и провести обработку данных. Вопрос лишь один: ЗАЧЕМ? Чтоб написать статью и получить лайки? Ну так и до психушки недолго.

Все кто зантмается аи сейчас — это чтоб ПРОДАТЬ свои наработки тем кто хочет это купить. Как я вижу целью всего этого аи будет готовое решение, которое можно продать конечному покупателю на завод или в офис. Стоить оно будет недорого.

По мне так тебе надо не трейдингом заниматься а идти работать на дядю. Дядя хотя бы направит твой труд в нужное русло а не в воздух

— Показать 4 ответа



Ещё 11 комментариев

Напишите комментарий...



ОТПРАВИТЬ

Читайте на SMART-LAB:

Откройте счёт в ВТБ Мои Инвестиции

RU000A100KH9	RUB	2,38	A+	16,94%	16,63%
RU000A100CK7	RUB	1,84	A-	21,34%	20,76%
Средний курсовый:					
RU000A100BZ7	USD	2,83	AA	7,98%	7,90%
RU000A100B06	USD	2,63	AA	8,60%	8,06%
RU000A100AP0	USD	1,74	AA	9,39%	9,15%

выпуски облигаций. Участие в первичном размещении может...



БКС Мир инвестиций

21.11.2025



RENI провела звонок для инвесторов и аналитиков по итогам раскрытия результатов за 9М 2025 года

Для удобства вопросы собраны в единый документ, размещенный на нашем сайте в разделе для инвесторов...



Ренессанс страхование

21.11.2025

Инфляция активов против инфляции потребительской – что действительно тормозит снижение ставки?

В публичном поле инфляция традиционно отождествляется с индексом потребительских цен (CPI). Именно этот показатель определяет траекторию ключевой ставки и служит номинальной...



Газпромбанк

20.11.2025

Установите приложение Смартлаба:



RuStore



AppGallery



App Store



О смартлабе

Реклама

Полная версия



Московская Биржа является спонсором ресурса smart-lab.ru
Источник: ПАО Московская Биржа

Откройте счёт в ВТБ Мои Инвестиции