

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: ['епренося', '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
=====
```

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	<a href="https://habr.com">https://habr.com</a>	113	2019-11	2025-09	0,0	2 192 876,0	4 497,0
3	<a href="https://pikabu.ru">https://pikabu.ru</a>	61	2024-09	2025-09	0,0	506 243,0	726,0
4	<a href="https://smart-lab.ru">https://smart-lab.ru</a>	50	2024-09	2025-09	0,0	453 955,0	1 998,0
5	<a href="https://www.youtube.com">https://www.youtube.com</a>	22	2013-02	2025-09	0,0	182 381,0	22,0
6	<a href="https://alenka.capital">https://alenka.capital</a>	1	2025-09	2025-09	0,0		0,0
7	<a href="https://fkvikings.timepad.ru">https://fkvikings.timepad.ru</a>	1	2025-09	2025-09	0,0		0,0
8	<a href="https://rationalanswer.club">https://rationalanswer.club</a>	7	2024-11	2025-08	0,0	45 078,0	39,0
9	<a href="https://vc.ru">https://vc.ru</a>	11	2019-11	2025-08	0,0	11 447,0	9,0
10	<a href="https://t.me">https://t.me</a>	12	2019-03	2024-12	0,0	79 200,0	122,0
11	<a href="https://www.1tv.ru">https://www.1tv.ru</a>	1	2024-12	2024-12	0,0		0,0
12	<a href="https://sprut.ai">https://sprut.ai</a>	3	2024-05	2024-11	0,0		0,0
13	<a href="https://t-j.ru">https://t-j.ru</a>	49	2020-04	2024-10	0,0	2 300 823,0	4 186,0
14	<a href="https://github.com">https://github.com</a>	10	2019-09	2023-06	0,0	0,0	0,0
15	<a href="https://podcast.ru">https://podcast.ru</a>	1	2022-11	2022-11	0,0		0,0
16	<a href="https://medium.com">https://medium.com</a>	1	2020-02	2020-02	0,0		0,0
17	<a href="https://zen.yandex.ru">https://zen.yandex.ru</a>	1	2019-12	2019-12	0,0		0,0
18	<a href="https://3dtoday.ru">https://3dtoday.ru</a>	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	<a href="https://z-wave.ru">https://z-wave.ru</a>	16	2014-11	2015-04	0,0		0,0

LynxReport: учёт публикаций  [Node.js] 

Недавно передо мной встал 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

Для этого цели у меня осталось всего один свободный 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 `comment-sentiment-analyzer` by `empenoso`. The repository is public and contains 6 commits. The commit history includes:

- `main` branch, 1 branch, 0 tags
- Commits:
  - `upd` (805ee0 · 6 minutes ago)
  - `habr_comments` (<https://habr.com/p/953034/>) (7 minutes ago)
  - `smart-lab_comments` (<https://habr.com/p/953034/>) (7 minutes ago)
  - `t-j_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)
  - `t-j_parser.py` (<https://habr.com/p/953034/>) (7 minutes ago)

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

**Readme**, **MIT license**, **Activity**, **0 stars**, **0 watching**, **0 forks**, **Report repository**.

**Releases**: No releases published.

**Packages**: No packages published.

**Languages**: Shell 62.7%, Python 33.2%, Dockerfile 4.1%.

**Analizator tonalnosti kommentariiev na Docker**

version 1.0, Python 3.9+, Docker Required, PyTorch 2.x

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



⌚ Недавние	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
★ Избранное	477920.json	480690.json	481566.json	481858.json	482800.json	485594.json	485848.json	486716.json	487436.json	489610.json
🏠 Домашняя папка	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
💻 Рабочий стол	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
📹 Видео	491182.json	494210.json	495264.json	495324.json	498268.json	503646.json	505674.json	506550.json	506720.json	515316.json
📄 Документы	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
📥 Загрузки	515732.json	533016.json	562546.json	577052.json	645935.json	680390.json	683188.json	694116.json	710584.json	712180.json
🖼 Изображения	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
🎵 Музыка	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
🗑 Корзина	721846.json	728840.json	733754.json	739168.json	742472.json	773620.json	777376.json	783066.json	804475.json	807961.json
+ Другие расположения	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
	811171.json	811681.json	817117.json	820249.json	825508.json	833648.json	835202.json	836450.json	836718.json	837616.json
	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
	838322.json	843298.json	843488.json	846938.json	849556.json	850992.json	855302.json	855742.json	855864.json	857402.json
	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
	860162.json	860706.json	863762.json	865558.json	867068.json	871234.json	871966.json	873242.json	874412.json	875846.json
	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
	876216.json	877264.json	879574.json	882608.json	883704.json	885308.json	887440.json	889932.json	891786.json	893006.json
	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
	896288.json	898174.json	900162.json	902334.json	903434.json	907030.json	907682.json	909852.json	912528.json	917420.json
	{}	{}	{}	{}	{}	{}	{}	{}	{}	{}
	919416.json	922148.json	925018.json	927238.json	929258.json	930340.json	933128.json	933806.json	935098.json	939390.json
	{}	{}	{}	{}	{}	{}	≡	habr_comme_ncts_positive_com..._ts.txt	+Хабр_выбо_рка.txt	
	942232.json	944670.json	945864.json	948224.json	948894.json	951288.json				

## 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`.

ПОЗИТИВНЫМ.

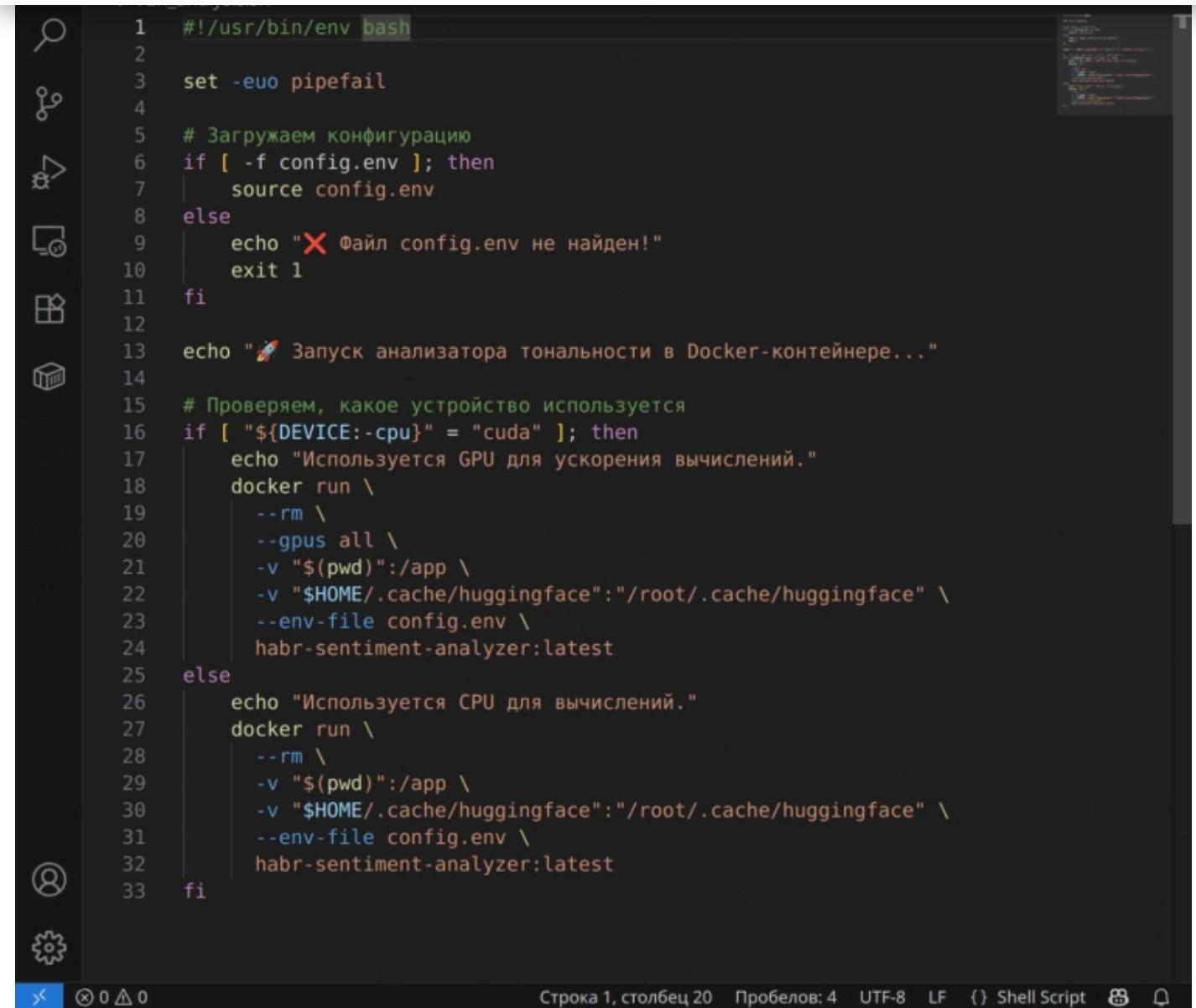
- EXCLUDE\_AUTHORS: список авторов, чьи комментарии нужно игнорировать (например, мои собственные ответы).
- MIN\_COMMENT\_LENGTH: отсеивает слишком короткие и неинформативные комментарии.

## 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.



```
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*.

```
=====  
💬 АНАЛИЗАТОР ТОНАЛЬНОСТИ КОММЕНТАРИЕВ  
=====  
  
⚙️ КОНФИГУРАЦИЯ  
=====  
📺 Модель: cointegrated/rubert-tiny-sentiment-balanced  
💻 Устройство: cpu  
📊 Порог позитивности: 0.8  
👤 Исключить авторов: ['емреносо', '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 комментариев в файл: habr_comments/habr_comments_positive_comments.txt  
📝 Сохраняю 182 комментариев 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.  
=====  
nike@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](https://t-j.ru/guide/excel-kotirovki/#c2857460)

Автор: DashBerlin

Дата: 2025-08-23T00:18:43.000Z

Текст: Впервые решил заглянуть в подобный обзор, порадовала позиция количества закладок, интересено, 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](https://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 pragmatичнее — 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 pragmatically и помнить, что иногда самое простое решение — самое верное.

7 октября 2025

**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

9.6K



☆ 9

44

32



## Михаил Шардин

Пермь

398 4 228

с 23 января 2019

+ Подписаться

44 КОММЕНТАРИЯ

Сначала старые ▼



ValiBakS

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 4 ответа

+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)

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

— Показать 2 ответа



Михаил Шардин ★

07 октября 2025, 08:35



Вазелин, вам же билет дают на конференцию — вот давайте там и пообщаемся? ;)



Ещё 12 комментариев

Напишите комментарий...



ОТПРАВИТЬ

## Читайте на SMART-LAB:



### Формат QSH: поддержка и особенности работы

Недавно OsEngine начал поддержку бинарного формата хранения и трансляции данных по стаканам. Это было нужно, чтобы...

OsEngine

18:01

Промышленная автоматизация — один из ключевых трендов 2026 в ИТ

SIO

### Промышленная автоматизация — один из ключевых трендов 2026 в ИТ #SOFL\_тренды

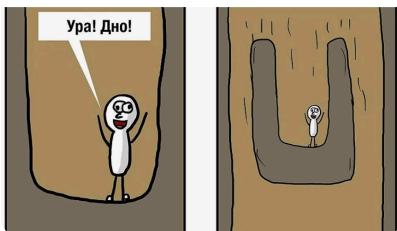
Сегодня промышленность все чаще смотрит на ИТ как на инструмент для наращивания мощностей. Для российской...



Гюокление ↗ относится к потреблению прагматичнее, чем остальные. Для них важны не громкие слова и статус, а понятна...

МГКЛ 10:00

кабре идет ко дну - хуже не было никогда



## Хэдхантер. Ситуация на рынке труда в декабре идет ко дну - хуже не было никогда

Вышла статистика рынка труда за декабрь 2025 года, которую Хэдхантер публикует ежемесячно, что же там интересного:...



Mozgovik

13.01.2026

Установите приложение Смартлаба:



RuStore



AppGallery



App Store



О смартлабе

Реклама

Полная версия



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