

The logo for ITMO, consisting of the letters 'I', 'T', 'M', and 'O' in a bold, white, sans-serif font. The 'I' has a small dot above it. The background is a dark purple grid with white wavy lines.

Научный Open Source в России: состояние и тенденции

Юрий Каминский
Андрей Гетманов
ITMO.OpenSource

О чём этот доклад?



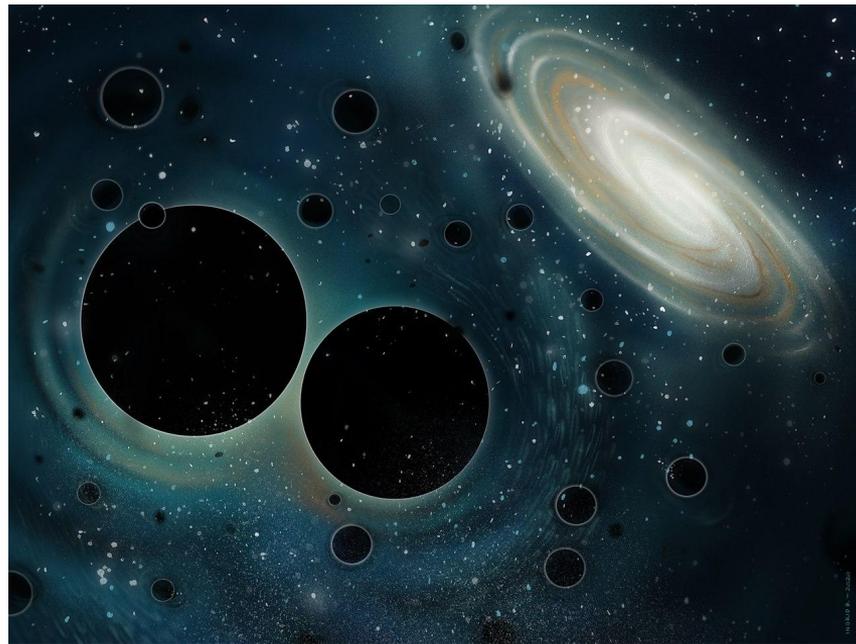
1. Почему написать статью уже недостаточно
2. Краткий анализ поля российских научных open source библиотек
3. Подборка лабораторий с разнообразными и интересными открытыми библиотеками

Good science requires good record keeping

Результаты события были опубликованы в 2020, а в 2021 были перепроверены (Brown et al., 2021)

Несмотря на изначально открытые данные и код, при воспроизведении возникли сложности: в оригинальной статье не была указана версия кода, не было скрипта для ключевых иллюстраций

Прوماхи встречаются в открытой науке даже такого уровня



GW190521: волновой всплеск вследствие слияния двух чёрных дыр

Почему важно открывать и поддерживать код?



Каждая хорошая работа над программным обеспечением начинается с реализации персональных целей разработчика

– Э. С. Реймонд, “Собор и базар”

Научное знание – это артефакт: статья, код к статье, фреймворк
Но поддерживаемая библиотека – более живучий, устойчивый
(holistic) артефакт (Liem et al., 2023)

Устойчивость улучшает:

- внешнюю воспроизводимость (Karoor et al., 2022)
- внутреннюю воспроизводимость (Barba et al., 2021)
- расширяемость

Последнее особо ценно, потому что способствует диалогу

Фреймворк – это не код к статье и не бенчмарк, а **продукт**:

- им можно пользоваться для разных научных целей
- его можно легко адаптировать под свои нужды
- он поддерживается и развивается

Ограничения:

- многие открытые фреймворки часто тяжело найти
- 6 ВУЗов* (ИТМО, МФТИ, ВШЭ, Сколтех, Иннополис, AIRI)
- 3 компании (Яндекс, Тинькофф, Сбер)
- Сводная оценка Library.io

Исследуемые репозитории: вузы



Организация	Описание	Самые популярные репозитории
ITMO AIM.CLUB	Объединенный репозиторий AI/ML фреймворков ITMO	FEDOT , BAMT , GEFEST , GOLEM
HSE LAMBDA	Лаборатория анализа больших данных ВШЭ	hsemotion , roerich , probaforms
МФТИ, SPC	Подборка проектов МФТИ	DeepPavlov , kmath
Сколтех	Новые официальные форки бывшего репозитория Сколтеха	ttoy , h2tools
AIRI	Репозитории с проектами Сбера	pogema , ai_toolbox , eco4cast

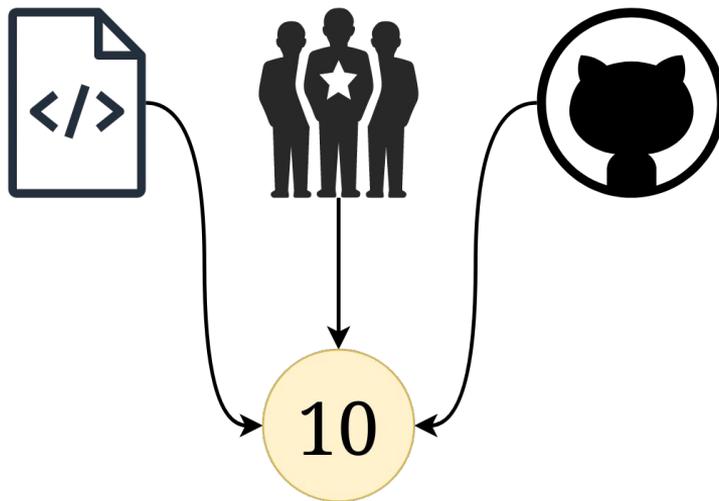
Исследуемые репозитории: компании



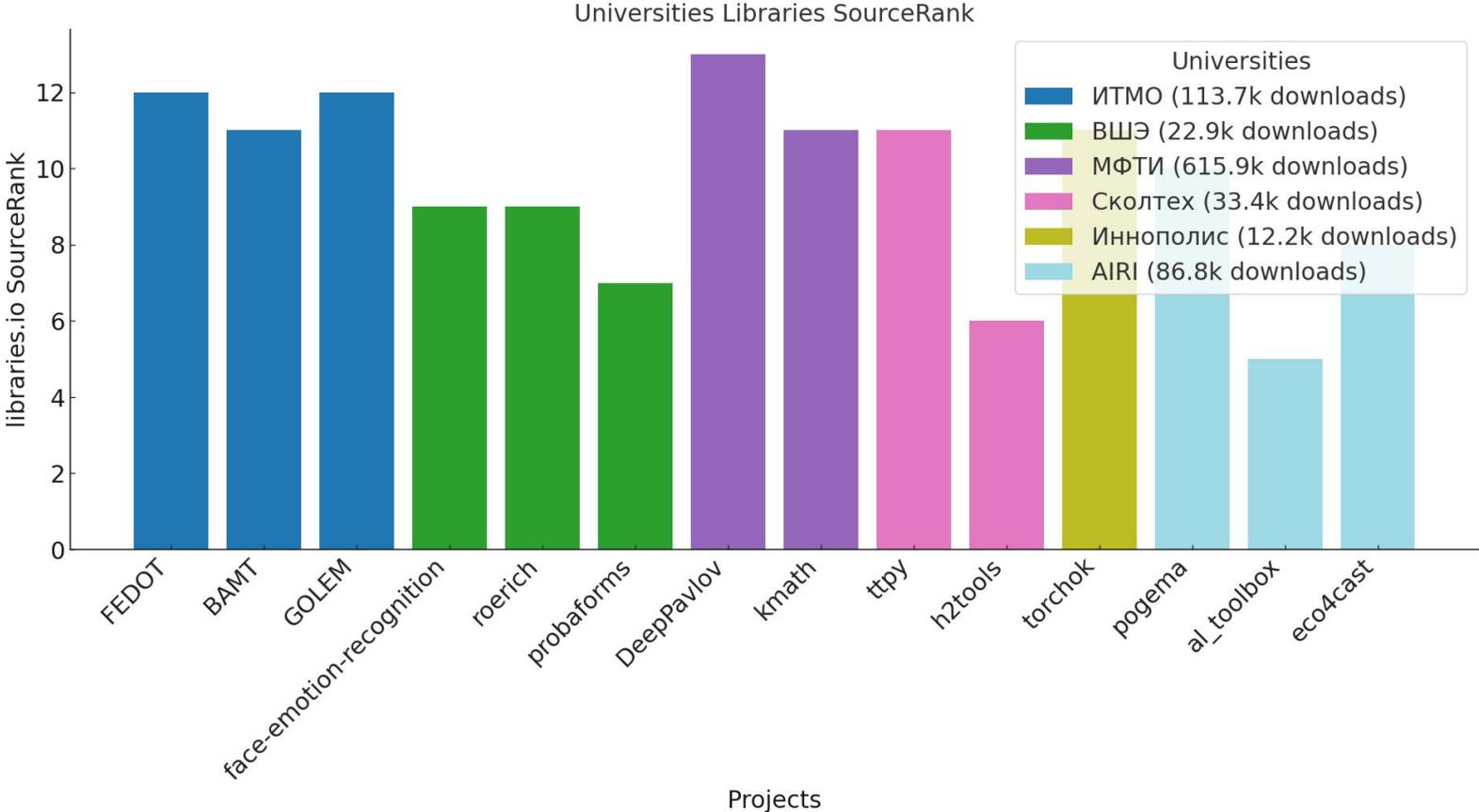
Организация	Описание	Самые популярные репозитории
Yandex	Open source продукты, зародившиеся в Яндексе	catboost , ytsaurus , clickhouse
ETNA-team, corl-team	Новые официальные форки бывшего репозитория Tinkoff	etna , corl , reBRAC
sb-ai-lab, ai-forever, AIRI	Репозитории с проектами Сбера	lightautoml , autowoe , eco2ai

Требования к методологии:

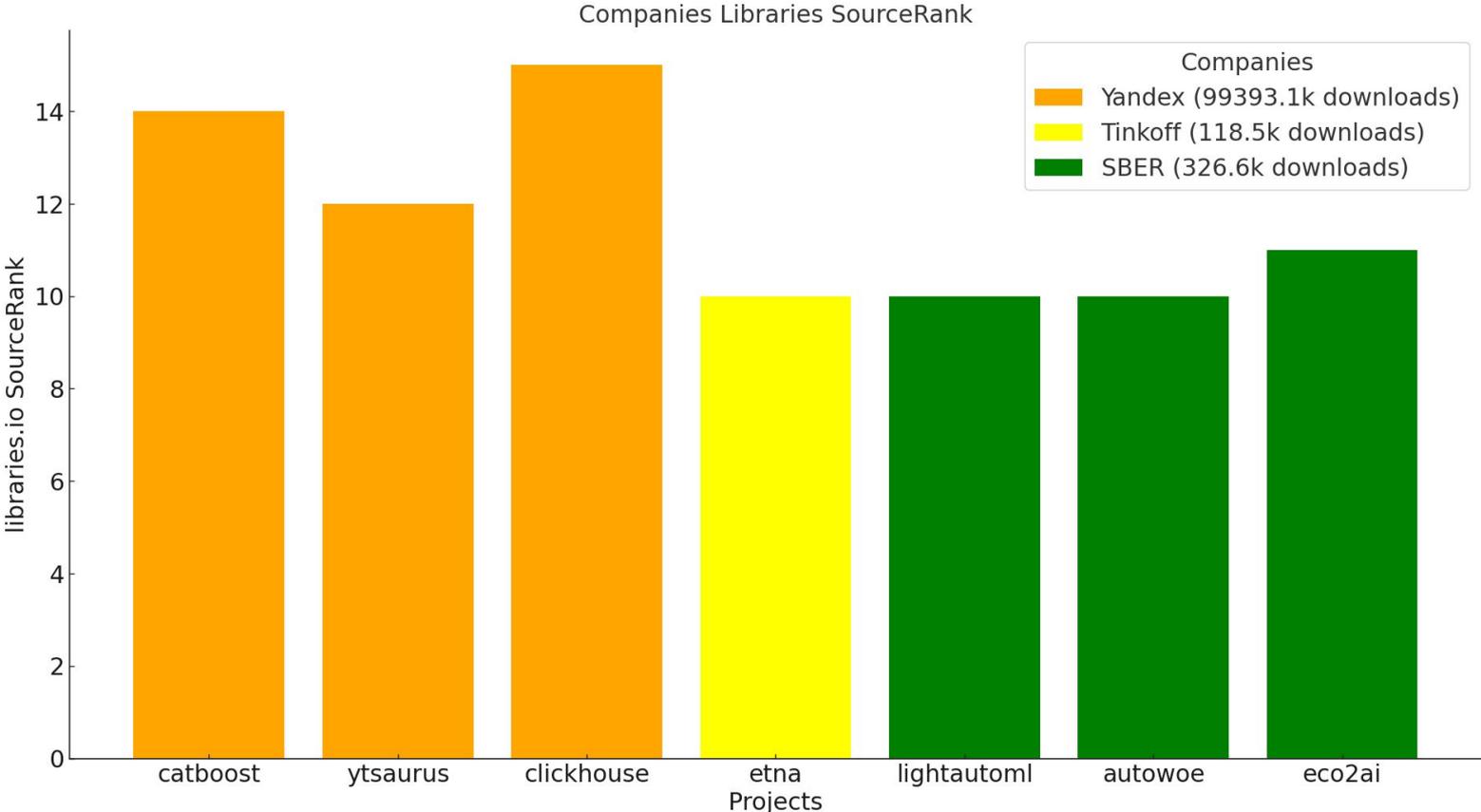
- Возможность сравнения мейнстримовых и “нишевых” инструментов
- Вывод общей оценки из большого количества параметров



Сводная оценка Library.io: университеты



Сводная оценка Library.io: корпорации



1. Open source в ML – не веяние моды, но критический запрос, который поможет структурировать галолирующий ресёрч
2. Оценка `libraries.io` хорошо выделяет заброшенные репозитории
3. “Нишевые” репозитории с маленьким количеством звезд могут конкурировать с мейнстримными
4. Такой интегральной оценке не хватает анализа кода

1. D. A. Brown, K. Vahi, M. Taufer, V. Welch and E. Deelman, "Reproducing GW150914: The First Observation of Gravitational Waves From a Binary Black Hole Merger," in Computing in Science & Engineering, vol. 23, no. 2, pp. 73-82, 1 March-April 2021, doi: 10.1109/MCSE.2021.3059232.
2. Liem, Cynthia CS, and Andrew M. Demetriou. "Treat societally impactful scientific insights as open-source software artifacts." 2023 IEEE/ACM 45th International Conference on Software Engineering: Software Engineering in Society (ICSE-SEIS). IEEE, 2023
3. Kapoor, Sayash, and Arvind Narayanan. "Leakage and the reproducibility crisis in ML-based science." arXiv preprint arXiv:2207.07048 (2022)
4. Barba, Lorena A. "Defining the role of open source software in research reproducibility." Computer 55.8 (2022): 40-48.

**Спасибо
за внимание!**

ITMO *re than a*
UNIVERSITY

Юрий Каминский
jkaminski@niuitmo.ru