

The logo for ITMO University, consisting of the letters 'ITMO' in a bold, white, sans-serif font. The background is a dark purple grid with white wavy lines on the right and bottom edges.

ITMO

Как и зачем привлекать контрибьюторов в открытые проекты

Никитин Николай, к.т.н,
руководитель направления открытого кода,
руководитель лаборатории автоматического машинного обучения,
ИЦ «Сильный ИИ в промышленности», ИТМО

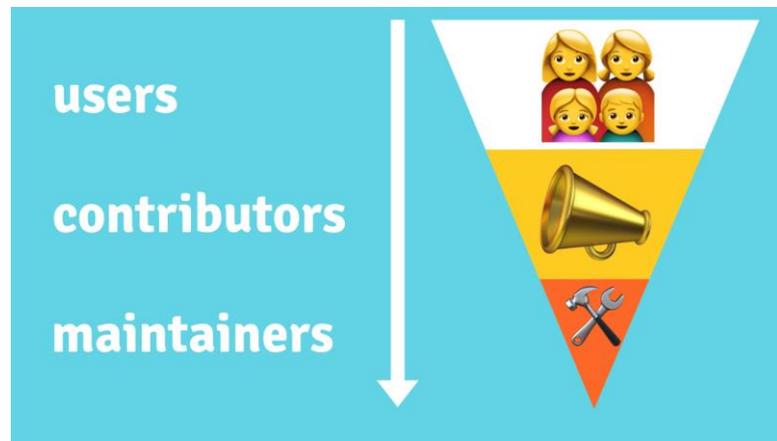
Санкт-Петербург – 03.10.2023

Кто такие контрибьюторы и зачем их искать?

Контрибьютор (contributor) – участник разработки открытого ПО в репозитории.

Он может быть «внутренним» – участником основной команды разработки или «внешним» – сторонним пользователем, который спонтанно решил принять участие в развитии репозитория.

Мы поговорим как раз о привлечение последних.



<https://mikemcquaid.com/the-open-source-contributor-funnel-why-people-dont-contribute-to-your-open-source-project/>

А нужны ли внешние участники вообще?

Для проектов:

- + Популяризация
- + Оперативный багфикс
- + Помощью в развитии, «обратная связь»
- Может быть проще сделать самому, через разобраться с чужим кодом

Для участников:

- + Опыт участия в проектах, «работа на резюме»
- + Нетворкинг, ценные знакомства
- + Влияние на развитие проектов
- Почти никогда не оплачивается

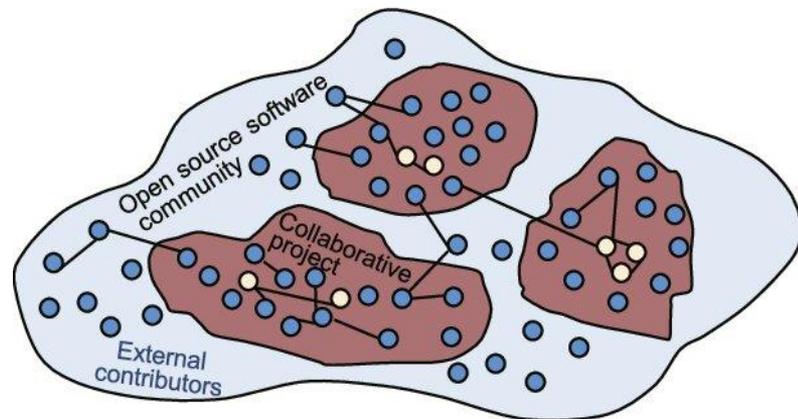


Схема опенсорс-сообщества

(no Li W. et al. Crowd intelligence in AI 2.0 era //Frontiers of Information Technology & Electronic Engineering. – 2017. – Т. 18. – С. 15-43.)

С чего начать контрибьюторам:

<https://github.com/firstcontributions/first-contributions>

А на что смотрят сами контрибьюторы?

1) Репозиторий жив?



2) Есть ли по проекту актуальные задачи?



3) Оперативно ли рецензируются изменения?



4) Интересен ли проект кому-то ещё?



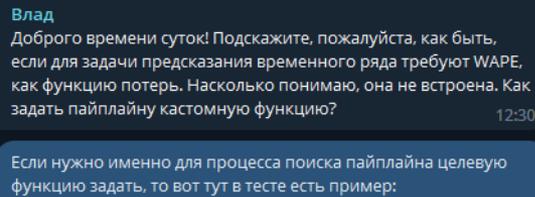
Used by 38



5) Понятен ли README? Есть ли тесты (хотя бы зачаточные)? Есть ли документация?



6) Есть ли где пользователю задать вопросы, если что-то непонятно (в чате или issue)?



7) Понятно, чем решение лучше аналогов (и знают ли авторы об этих аналогах вообще?)

Плохие практики авторов библиотек

1) Игнорировать внешние issue

BUG : Spearman correlation error generated from missing otypes in np.vectorize #61 opened on Jun 16 by Andralonescu
ADD : may you add code examples fro each algorithm #60 opened on May 22 by Sandy4321
BUG : #57 opened on Sep 4, 2022 by LiWei-DU
BUG : Feature selection example from documentation not working #52 opened on Jun 30, 2021 by nicl-nno
Allow installing with conda #45 opened on Feb 5, 2021 by ch3rn0v
TypeError #43 opened on Jan 21, 2021 by hugoas
Different results #42 opened on Jan 17, 2021 by yusufmet

Итог: библиотека заброшена

2) Игнорировать pull request-ы

Make everything work again #6

Open andreygetmanov wants to merge 1 commit into `kiok46:master` from `andreygetmanov:pr_fix_agetmanov`

Conversation 0 Commits 1 Checks 0 Files changed 4

andreygetmanov commented on Aug 3

- switched off parallel threading due to the logics change
- changed geolocator to geopy because Google Maps need access token
- updated countries.js because some geocodes weren't identified on the web map
- added auth by GitHub token due to the 10 req/hour unath limit
- fixed login parser due to new GitHub web design
- updated requirements.txt
- minor PEP-8 changes

1

Make everything work again

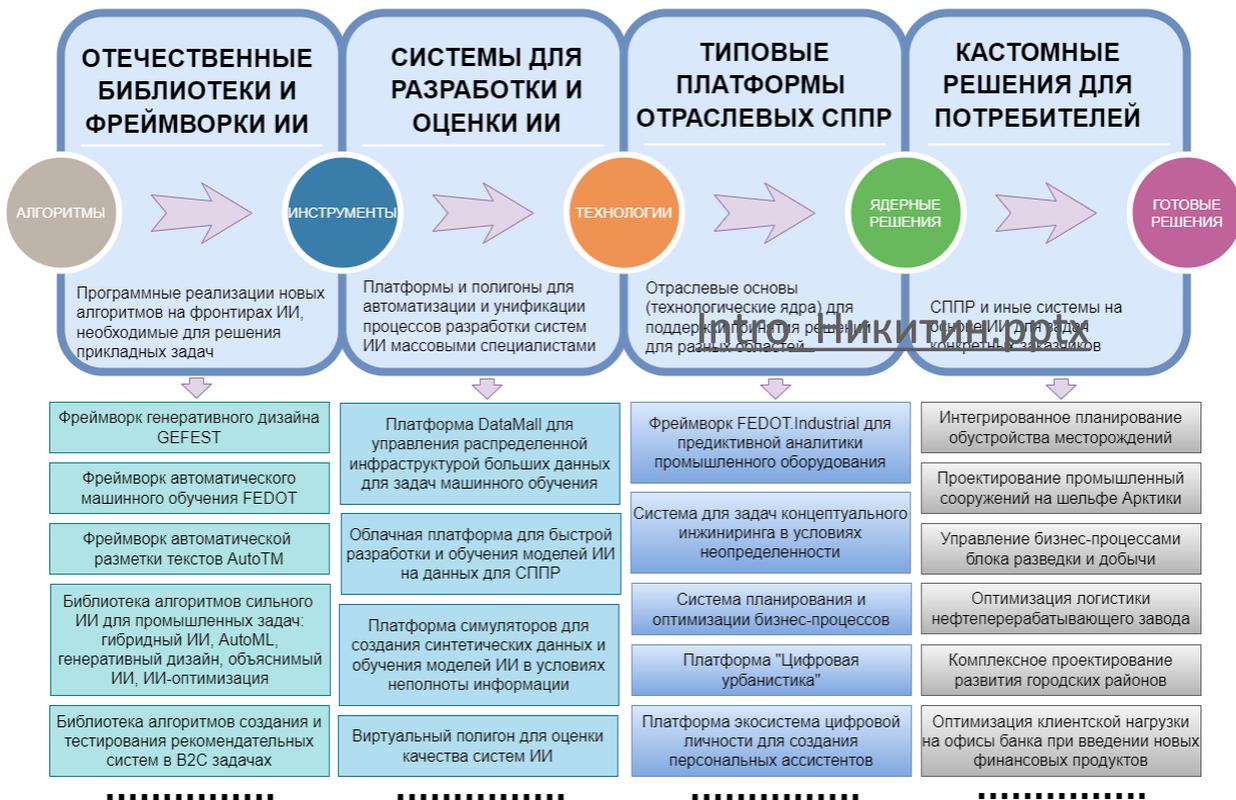
andreygetmanov mentioned this pull request on Aug 3

Some things didn't work, but I fixed them #5 Open

Итог: рабочая версия ушла в fork

Продукты и разработки AIM

ИТМО



30+ НОВЫХ реализаций алгоритмов ИИ

6 инструментальных платформ для разработки систем ИИ

60+ решений в интересах конкретных заказчиков

Открытый код ИТМО в области ИИ



Pinned

FEDOT Public Automated modeling and machine learning framework FEDOT ● Python ☆ 553 🍷 77	BAMT Public Repository of a data modeling and analysis tool based on Bayesian networks ● Python ☆ 85 🍷 13
Fedot.Industrial Public Python framework for automated time series classification, object recognition and anomaly detection ● Python ☆ 56 🍷 3	GOLEM Public Graph Optimiser for Learning and Evolution of Models ● Python ☆ 45 🍷 4
GEFEST Public Toolbox for the generative design of geometrically-encoded physical objects using numerical modelling and evolutionary optimization ● Python ☆ 48 🍷 8	iOpt Public Framework of intelligent optimization methods iOpt ● Python ☆ 36 🍷 14

Repositories

🔍 Find a repository... Type ▾ Language ▾ Sort ▾

Фреймворков
библиотеки и
фреймворков в
области ИИ,
готовые к
использованию,
снабженные
тестами и
документацией.

<https://github.com/aimclub> и не только

Открытость и воспроизводимость в науке



Конференции, журналы и статьи тоже могут быть открытыми – добавление кода и данных повышает ценность публикации в разы.

Пример ресурса про «статьи с кодом»:

<https://paperswithcode.com/>

Наш пример: для материалов конференции International Young Scientists Conference 2023 создан отдельный репозиторий.

Deep Learning and Data-Driven Modelling

Section Papers 13 Papers with Open Code 9

ID	Title	Links
06	Light-weight ensembling of deep neural models for object recognition in remote sensing data	Code Data
15	Convolutional Neural Network Graph-based Embedding for Neural Architecture Search	Code
16	FAMLINN: Representation for Storing Neural Network Architecture	Code
18	Multimodal prediction of profanity based on speech analysis	Code
24	Proactive selection of machine learning models for small sample sizes based on PAC-learning theory	Code
31	CIS Multilingual License Plate Detection and Recognition Based on Convolutional and Transformer Neural Networks	Code Data
32	Predicting dataset size for neural network fine-tuning with a given quality in object detection task	Code
35	Forecasting of Sea Ice Concentration using CNN, PDE discovery and Bayesian Networks	Code
45	mHAR: a novel convolutional recurrent model for recognizing motion-based human activity	Code Data

<https://github.com/itmo-ai/YSC-2023-Papers>

Сообщество open-source в ИТМО

ИТМО

- Чат разработчиков и пользователей открытого научного кода - https://t.me/itmo_opensource.
Помощь в решении проблем с созданием открытых решений, анонсы мероприятий и полезных материалов.
- Студенческий клуб ИТМО.OpenSource
- Митапы «Научный опенсорс» - очные и онлайн-встречи сообщества



Репозиторий ITMO.OpenSource

Репозиторий с руководствами, лучшими практиками и шаблонами для создания открытых проектов:

<https://github.com/aimclub/open-source-ops>



Open-source-ops (by NSS Lab)



Этот репозиторий создан для хранения различных инструментов, скриптов, инструкций и руководств, которые могут быть полезны при создании проектов с открытым исходным кодом. Все материалы доступны под лицензией BSD-3.

Основные разделы

Инструкции

- С чего начать разработку open-source библиотеки;
- Зеркалирование GitHub -> GitLab;
- Мультиязычные README;
- Создание документации;
- Настройка ботов для репозитория.

Шаблоны

- Типовой шаблон README для open-source проектов (Версия в RST формате).

Лучшие практики и примеры

- Организация управления open-source проектом;
- Полезные ссылки для авторов open-source библиотек;
- Советы по работе в Pull Request-ax;
- Советы по популяризации репозитория.

Открытый код и наука

- Где опубликовать научную статью про OS-разработку?.

Открытый код в ITMO и не только

Как поддерживать открытые проекты

- Ставить звездочки. Больше звездочек – больше кредит доверия.
- Цитировать в научных статьях. Больше цитирований – выше вероятность, что статью и связанный с ней проект заметят.
- Лайкать в социальных сетях.
- Упомянуть в обзорах, подборках и т.п.
- Пробовать пользоваться и давать обратную связь.

Что делать, если заинтересовался

- У нас есть программы магистратуры: «ИИ в промышленности», «Большие данные и машинное обучение».
- Сообщество ITMO.Opensource – хорошая платформа для апробации полезных пет-проектов и участие в профессиональном сообществе.
- Контрибьютить во все наши открытые проекты можно свободно.
- Открыты для любого научного сотрудничества.

**Спасибо
за внимание!**

ITMO *re than a*
UNIVERSITY

nnikitin@itmo.ru