

Яндекс

# Разработка системы инкрементальных вычислений

Студент: Федор Бочаров  
Руководитель: Александр Юрченко

# Мотивация

Построение поискового индекса – batch-процесс, а значит он обладает огромным недостатком – *латентностью*. Это напрямую влияет на свежесть индекса – один из ключевых показателей поисковых систем.

# Что делать?

В 2010 году компания Google опубликовала статью про модель инкрементальных вычислений и систему Percolator, которая позволила сократить время обновления индекса в десятки раз.

Ключевая особенность – запуск кода, перестраивающего индекс, только **на обновленной** части индекса.

# Поставленные задачи

- Разобраться с существующей кодовой базой
- Разработать подсистему запуска пользовательского кода (триггеров) на произвольном ЯП
- Разработать инструментарий для сбора статистики
- Реализовать инкрементальный экспорт данных во внешние системы

# Запуск пользовательского кода

Что было:

- Наколеночный запуск python-кода
- Возможность запуска C++-кода при условии статической линковки ☹️

# Запуск пользовательского кода

Как стало:

- Единый, с точки зрения разработчика, способ запуска триггеров на любом языке программирования: каждый триггер имеет свой «хост», который знает, как его запустить и как передать ему данные
- Реализованы `shell-host` (запуск триггеров в дочернем `shell`) и `C++-host` (загрузка триггеров из `.so`-библиотек налету)

# Инкрементальный экспорт

Что было:

- Ничего

Стало:

- Пользователь указывает, какие данные как часто и куда экспортировать, а так же, как они должны быть интерпретированы (plain text, json, ...), система просто делает это

# Итоги

Пока подводить рано 😊

Промежуточные:

- Попробовал полный спектр девелоперских обязанностей: написание кода, исправление чужих багов, запуск в бета-тестирование, поддержка пользователей
- Получил опыт работы с распределенными системами

**Спасибо за внимание**