



АЛГОРИТМ КОНТЕКСТНО-ЗАВИСИМОГО АННОТИРОВАНИЯ

Александр Салтыков

руководитель отдела проектирования

Андрей Канунников

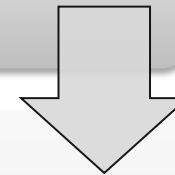
ведущий аналитик отдела проектирования

- R&D подразделение компании Ingate
- Специализация – создание массовых сервисов для интернет рекламы
- ROOKEE – первый продукт, SEO

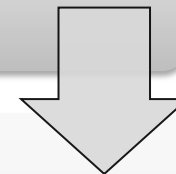
Цель разработки

- Автоматизация создания рекламных материалов
- Отработка технологий привлечения трафика
- Получить опыт участия в РОМИП

Парсер документов



Кластеризатор
пассажей



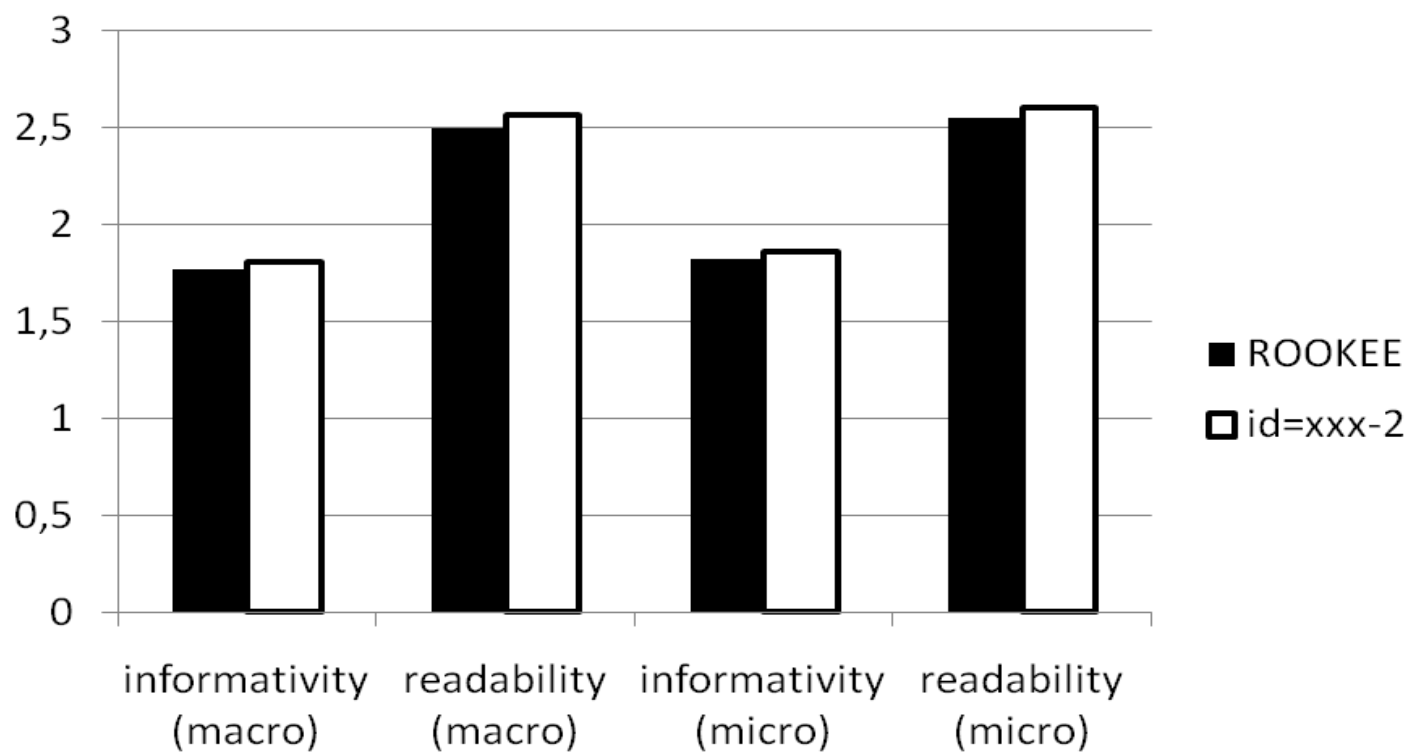
Конструктор
аннотаций

- Преобразование HTML в XML
- Предварительная редукция структуры XML-документа
- Редукция структуры документа по индикаторам
- Коррекция пунктуации

- Выделение пассажиров
- Кластеризация пассажиров
- Ранжирование кластеров

- Первичная обработка кластера
- Построение каркаса аннотации
- Формирование аннотации
- Контроль размера

Результаты участников



- Корректное выделение пассажей
- Эффективная кластеризации и ранжирование
- Полнота баз слов-переходов и транслитераций

≈ 52 %

в аннотации текст **Я**ндекса



- Разработан эффективный метод контекстно-зависимого аннотирования
- Метод успешно прошёл цикл внутренних испытаний в компании
- Высокая эффективность метода подтверждена экспертами РОМИП



Спасибо за внимание!
Вопросы?