

ARTM Online

Технология интерактивной визуализации тематических моделей

Дмитрий Федоряка

Московский физико-технический институт

25 января 2017 г.

- Инструмент для исследователей BigARTM
- Сайт для рекламы наших возможностей
- Среда для сбора оценок ассесоров

- Документ
- Тема
- Термин
- Иерархия
- Темпоральная визуализация

- Универсальность и простота.
- Простое добавление модулей (визуализация, именование тем).
- Визуальная привлекательность (d3).
- Скорость и масштабируемость (БД, AJAX).

Последовательность построения модели

- Коллекция \rightarrow мешок слов
- Мешок слов \rightarrow батчи ARTM
- батчи ARTM \rightarrow Построение модели $\rightarrow \Phi, \Theta, \Psi$
- $\Phi, \Theta, \Psi \rightarrow$ дерево тем и документов.
- Автоматическое именование тем.
- Визуализация.

- Веб-сервис на python и django с базой данных.
- Все объекты (документы, темы, термины, отношения) храним в базе данных
- Текст документов – в файлах.
- Есть интерфейс построения моделей, но можно их строить отдельно и загружать.
- Визуализации - как отдельные модули (2 файла: py, js).

- Упорядочивание тем (в процессе)
- Именованние тем
- Интерфейс ассесора (?)