

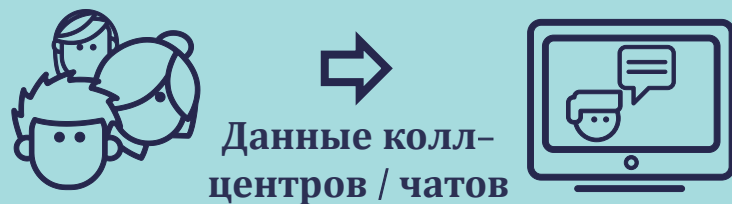
Обработка неструктурированных данных: SAS Visual Text Analytics

Дмитрий Каютенко

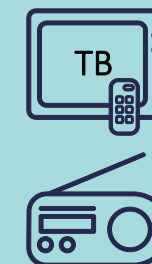
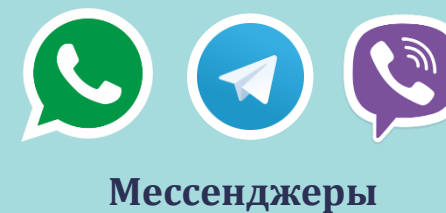
Консультант, Дирекция аналитических и промышленных
решений, SAS Россия / СНГ

Какие бывают текстовые данные?

Внутренние данные КОМПАНИИ

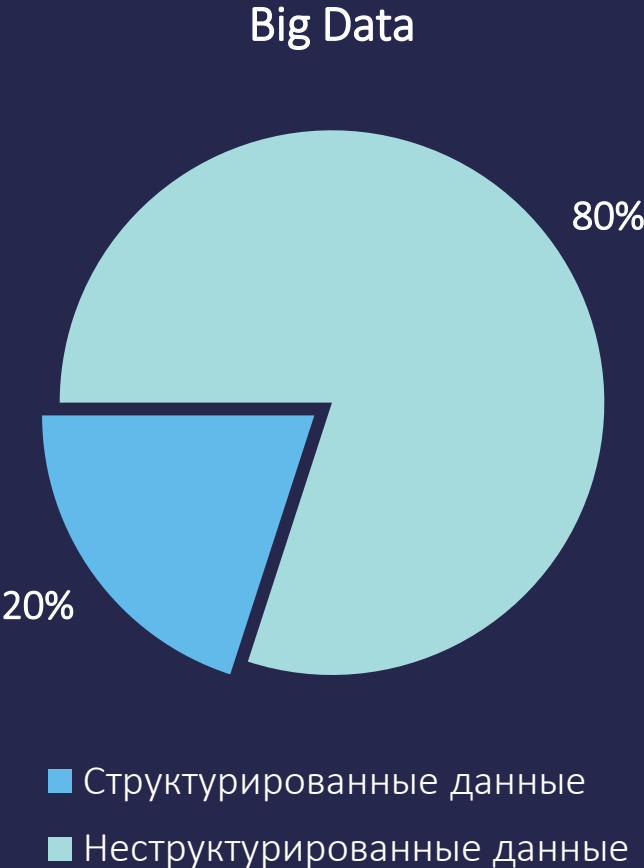


Внешние данные

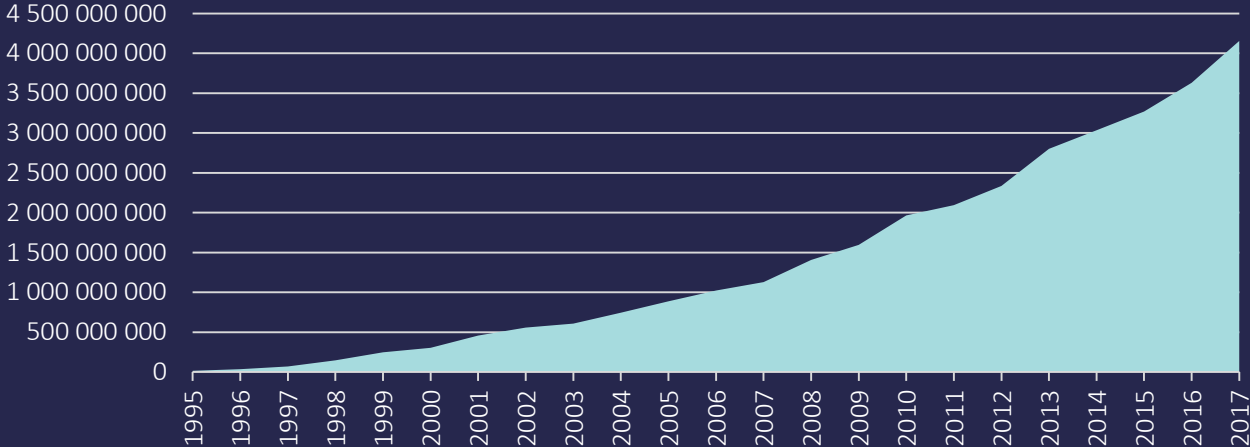


Почему важно уметь осуществлять автоматический анализ текста?

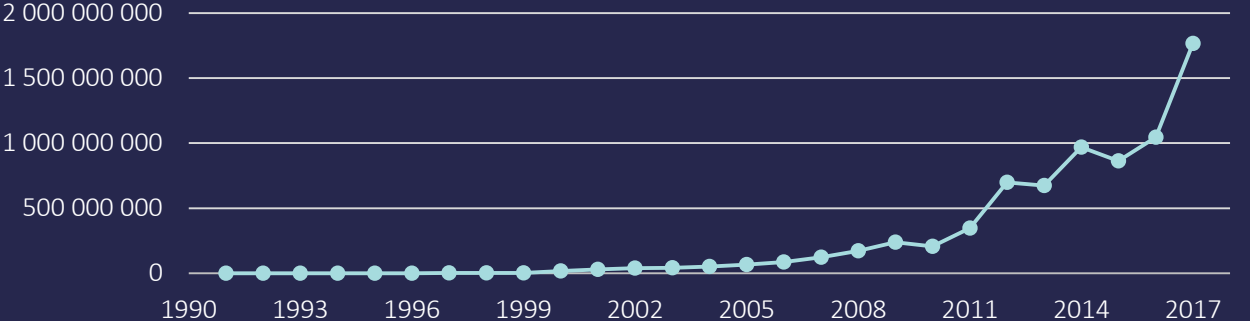
Большая часть «Big Data» – неструктурированный текст и его все больше



Роста числа пользователей Интернета



Роста числа вебсайтов в Интернете



Ключевые направления применения текстовой аналитики



Какие проблемы могут возникнуть при анализе текста?



- Отсутствие обучающих выборок



- Большие затраты на формирование обучающей выборки для автоматического построения модели



- Большие временные затраты на ручную настройку модели



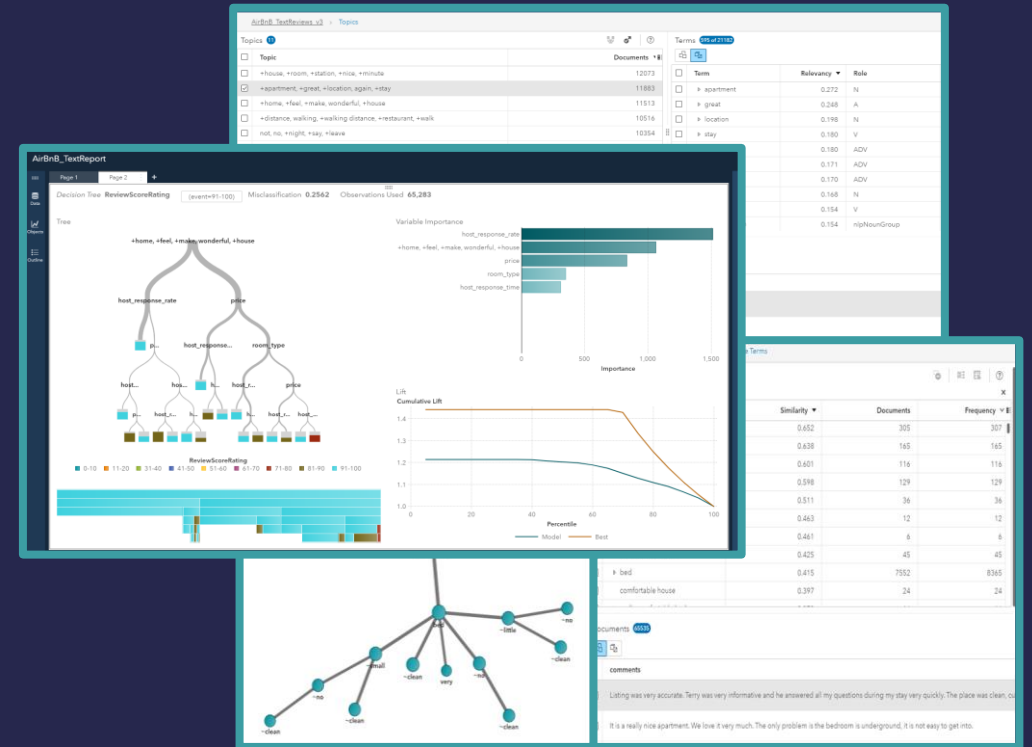
- Ограниченность точности статистических моделей качеством данных

SAS Visual Text Analytics

Единый инструмент для решения любых задач, связанных с анализом текста

Ключевые возможности:

- Извлечение информации из неструктурированных текстовых данных
 - Автоматическое выявление тематик
 - Категоризация
 - Выявление упоминаний объектов (персон, компаний и т.п.)
 - Выявление фактов (кто, что, когда и с кем сделал)
- Возможность сочетания алгоритмов Natural Language Processing, машинного обучения и экспертных знаний о предметной области



Демонстрация

SAS Visual Text Analytics

Спасибо за внимание!