Brand Analytics

Вы (жить) в тренде

Анализ миллиардных социопотоков

http://br-analytics.ru

Изменение потоков информации за последние 20 лет

Что, где, когда?

«… Я живу в мире, который кто-то придумал, не затруднившись объяснить его мне, а, может быть, и себе. Тоска по пониманию,-— вдруг подумал Перец.— Вот чем я болен — тоской по пониманию.»

Братья Стругацкие, «Улитка на склоне», 1965

Изменение потоков информации за последние 20 лет

- **Статьи (СМИ)**
  Factiva, Lexus-Nexus, Медиалогия
- **Документы (онлайн-СМИ, блоги)**
  Google, Яндекс.Блоги и т.д.
- **Мнения (соцсети-общение)**
  Brand Analytics, DataSift, Gnip
- **Сообщения и открытые данные**
  (люди и системы) — ?

График на иллюстрации логарифмический,
в реальной пропорции кривой не видно,
она превращается в «вертикальный взлет»

Потоки превратились в океан информации

Человечество генерит в сутки 30+ миллиардов сообщений, из которых порядка 10% – публичных

- Как в океане информации найти главное течение, свой Гольфстрим?
- Как распознать зарождение информационного цунами?
- Кто является основными объектами интереса в инфополе?
- Как узнать «погоду» на завтра?
- Как построить тысячу сегментов аудитории и понять молодежь?

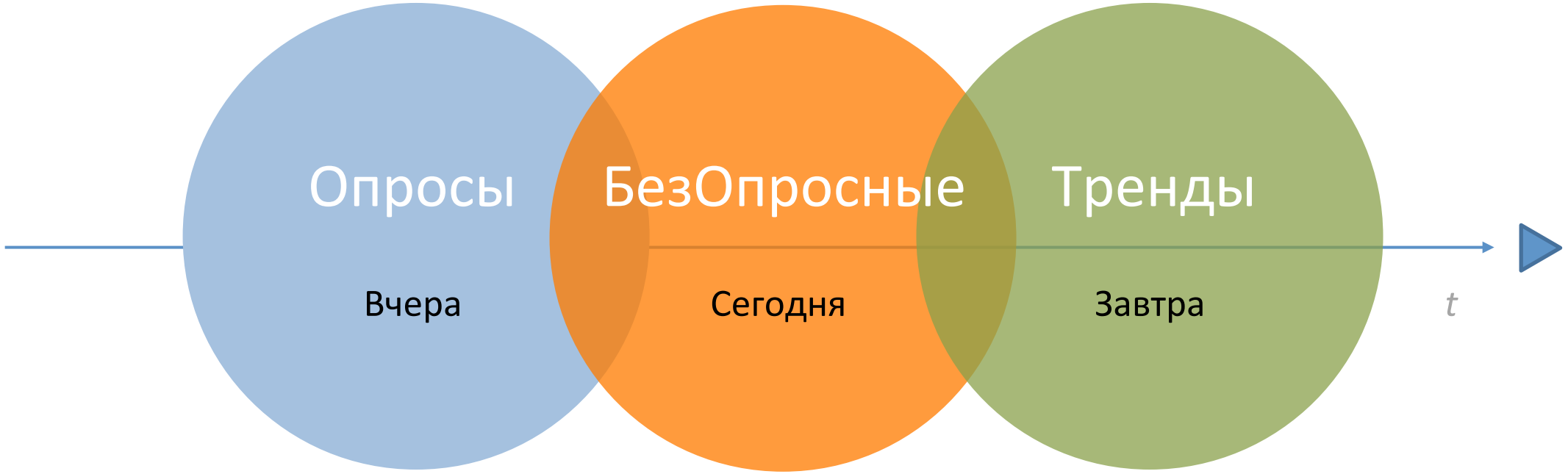
Развитие технологий исследований

Что делать?

«Выигрывает вовсе не тот, кто умеет играть по всем правилам; выигрывает тот, кто умеет отказаться в нужный момент от всех правил, навязать игре свои правила, неизвестные противнику, а когда понадобится — отказаться и от них.»

Братья Стругацкие, «Град обреченный», 1975

Развитие технологий исследований



Решение: тренд-системы

Эти и многие другие вопросы и задачи помогают решать современные технологии анализа высокоскоростных потоков социо-данных:

- **Что сейчас:** какие тренды и сюжеты являются основными в динамическом информационном поле индустрии, региона, страны и мира; какие структуры и персоны являются объектами внимания медиа и общества.

- **Что новое:** оперативное выявление новых трендов и сущностей.

- **Что будет:** прогностика — «вероятностный» мониторинг: выборы, развитие социума, влияние новых технологий и пр.

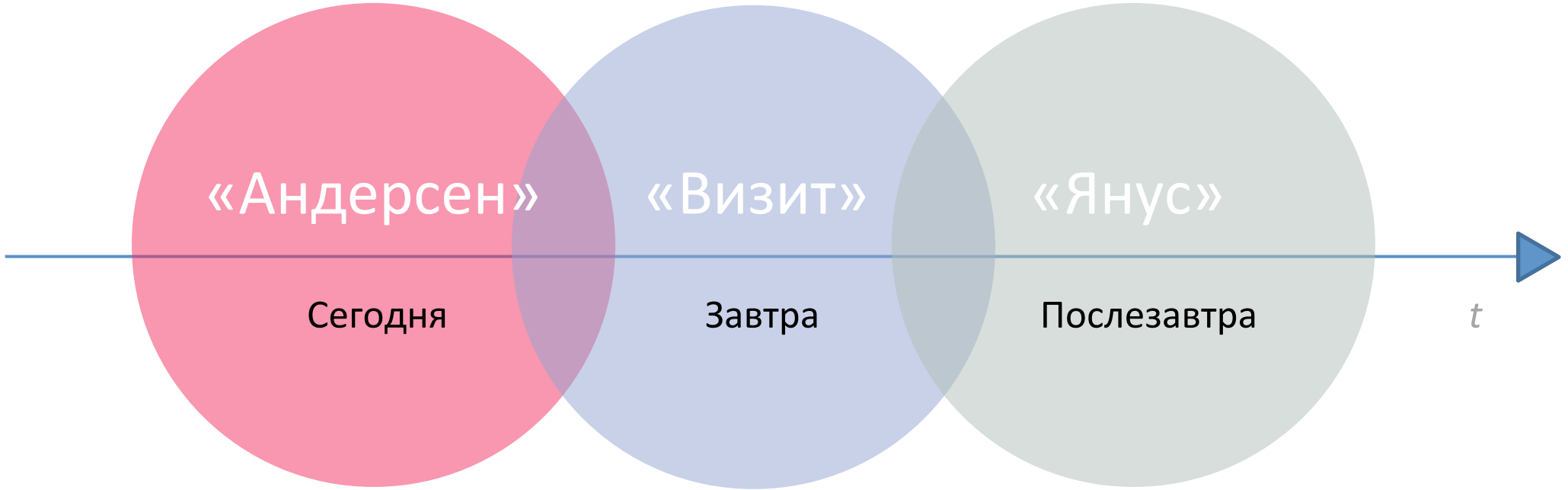
Тренд-системы Nooshere Technologies

Проект	Целевая аудитория	Зачем
«Андерсен» (основные сюжеты и объекты)	СМИ, PR/маркетинг и др.	Вовремя знать главное по своей тематике
«Визит» (зарождающиеся тренды)	Информационные агентства, Ситуационные центры, Инвест-фонды и др.	Выявление сигналов социума на начальной фазе роста
«Янус» (прогностика)	Основные игроки рынка, Госструктуры	Влиять на будущее, зная наиболее оптимальное место приложения усилий

NOOS.TECH

Тренд-системы Nooshere Technologies

NOOS.TECH



«Андерсен»: Оперативное выявление информационных трендов

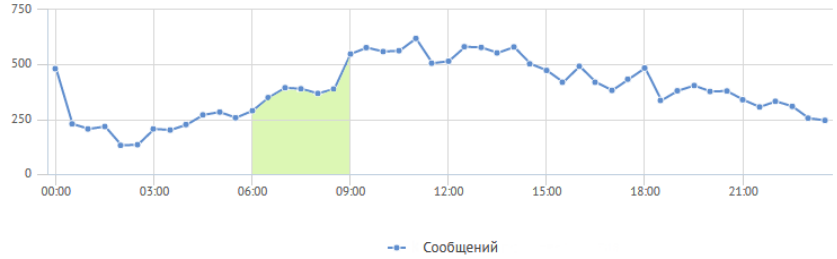
Тема: Нефть Тематика: Экономика Язык: Русский 10.04.2017 06:00 – 09:00

158

Немцы предложили России вложить $300 млрд в отказ от нефти и газа #lifenews #новости #new
URL: http://twitter.com/ru_newsmix/status/851307085665103874
Сообщений: 73
Объекты: Россия

Дворкович: Россия ожидала, что цены на нефть будут $55-60 после соглашения с ОПЕК, но этого не случилось
URL: http://vk.com/wall219700659_13134
Сообщений: 54
Объекты: Россия, ОПЕК, Дворкович

@proro2012: Россия может выйти из сделки с ОПЕК, что приведет к обвалу цен на нефть
URL: http://twitter.com/janna_com/status/851308044810153986
Сообщений: 31
Объекты: Россия, ОПЕК

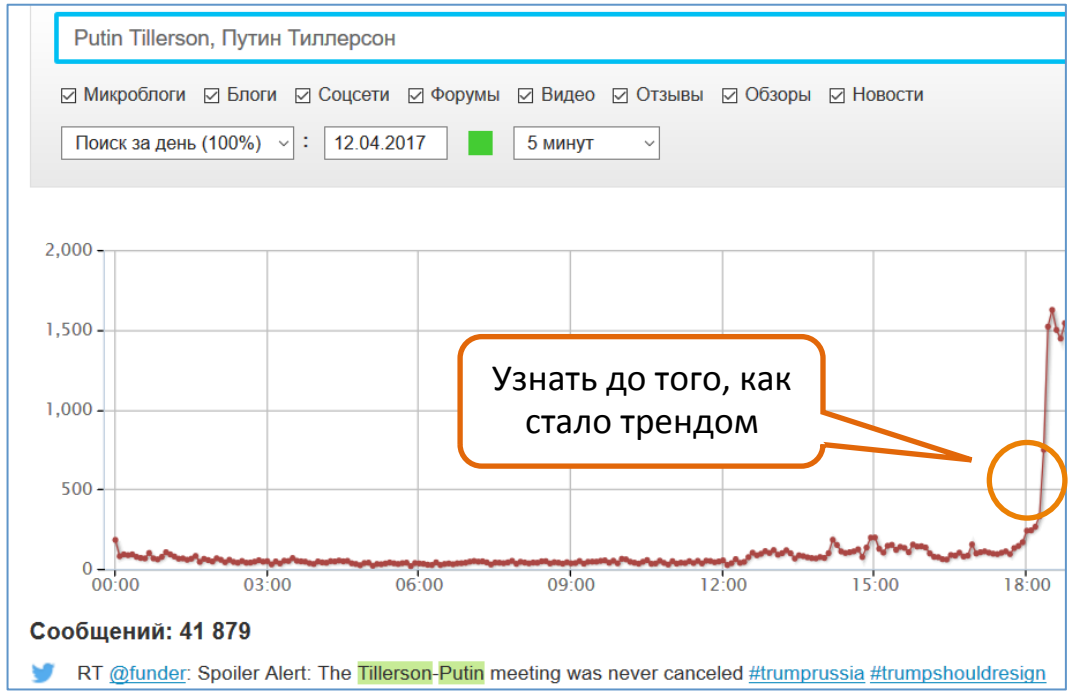


107

Росгеология оценит перспективы добычи нефти и газа в Якутии и Сибири: Оценкой недр Якутии и Сибири займётся... #обзор
URL: http://twitter.com/eleonorayakush1/status/851277837528989696
Сообщений: 62
Объекты: Росгеология, Якутия, Сибирь

«Андерсен» – технологии

«Визит»: Выявление зарождающихся трендов



На рисунке: Зарождение информационного тренда о встрече Путина и Тиллерсона

Знать о проблемах и возможностях в числе первых 5% !

Примеры:

1. Неожиданная встреча Путина и Тиллерсона.

Результат: Факт встречи сыграл резко на бирже против $ (защитный актив) - раз всё хорошо (встреча), значит можно оставаться и наращивать российские бумаги и рубль.

2. Инцидент с United Airlines - избиение пассажира и реакция в соцсетях.

Результат: Потеря капитализации $600 млн.

«Визит» – технологии

Платформа потоков данных

5-10 тысяч сообщений в секунду

Тематическая выборка

1/5/15/60 минут

Очистка данных,

лемматизация /стемминг

Расчет частоты, сравнение с базой ЧМС

(частотка на миллион слов)

Выявление тренд-слов,

формирование кластеров вокруг Т-слов

Лингво-модули:

кластеризация,

классификации,

NER

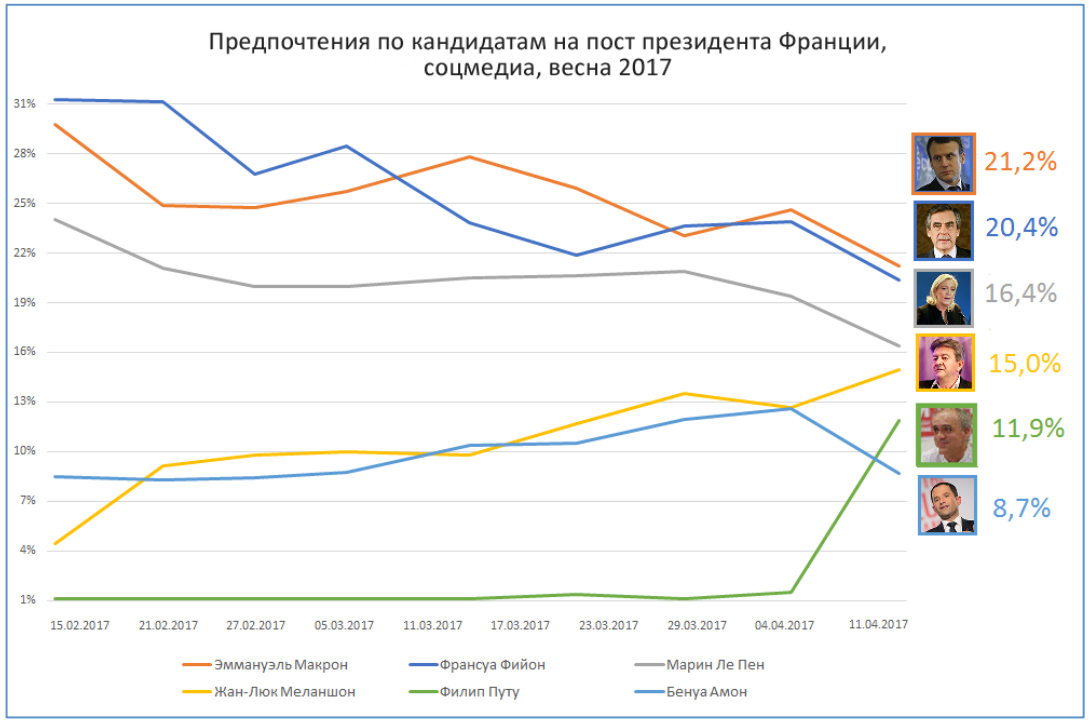
VIO

Пост-обработка:

группировка,

отправка сигналов и отчетов

«Янус»: Вероятностное будущее



На рисунке: Динамика изменения предпочтений по кандидатам на пост президента Франции, измеряемое на основе анализа потока сообщений французской аудитории социальных сетей. 20+ млн сообщений, 3+ млн авторов.

Взгляд назад – там застывшее прошлое, легко алгоритмизируемое. Широкое использование методов машинного обучения – это "продолжение прошлого в будущее". Работает, но не всегда, а ошибки – катастрофичны: Lehman Br., Nokia, Yahoo...

Взгляд вперед – варианты развития.
Конечные, например выборы, можно использовать локально:
Победа Трампа = рост золота, снижение песо.
Индустриальные (глобальные): Tesla, бумажная пресса, чат-боты.

Изменения "настроений" можно и нужно фиксировать постоянно и в ДИНАМИКЕ:

- Дни, недели, месяцы;
- Миллионы людей: пол, возраст, гео, тональность, влиятельность, виральность и т.д.;
- Миллиарды мнений и высказываний;
- Выявление болевых точек и точек роста;
- Формирование вероятностных стратегий. Всех возможных;
- Динамическая перестройка вероятностей.

Подводные камни

В реализации технологий – множество «подводных камней»:

- Бот-сети. Выявление и пополнение базы. Использование задействования — как сигнал интереса.
- Кластеризация текстовых документов не только на основе сходства,
  но и на основе семантической смежности или ассоциативности.

  Например,
  - В Лондоне передумали объявлять России новую холодную войну
  - Борис Джонсон: Запад не находится в состоянии новой холодной войны с РФ
  - В британском МИД заявили, что Запад не хочет новой холодной войны с Россией

  Все три примера являются одной новостью, тем не менее, лексика используется разная.
- Высокоскоростная лингвистика – использование только статистических методов
  не дает достаточной точности результатов.
- Высокие стоимости реализации, поскольку запредельные скорости потоков данных.

  Для примера несколько цифр:
  - Стоимость платформ сбора – Apple купила TopSy за $200 млн, а Twitter Gnip за $230 млн.
  - Стоимость высокоскоростной лингвистики – IBM купила AlchemyAPI за $100 млн
  - Стоимость аналитических систем – Salesforces купила Radian6 за $380 млн, HP Autonomy за $12 млрд

Brand Analytics

БРЕНД МОНИТОРИНГ И СОЦИАЛЬНАЯ АНАЛИТИКА

Спасибо за внимание!

http://br-analytics.ru | тел.: +7 (495) 105-95-01