

ОПРЕДЕЛЕНИЕ ТРЕНДОВ В СОЦ СЕТЯХ В СФЕРЕ КОСМЕТИКИ

Докладчик: Дарина Дементьева

КАК ВСЕ НАЧИНАЛОСЬ

Что популярно в
уходе за кожей?



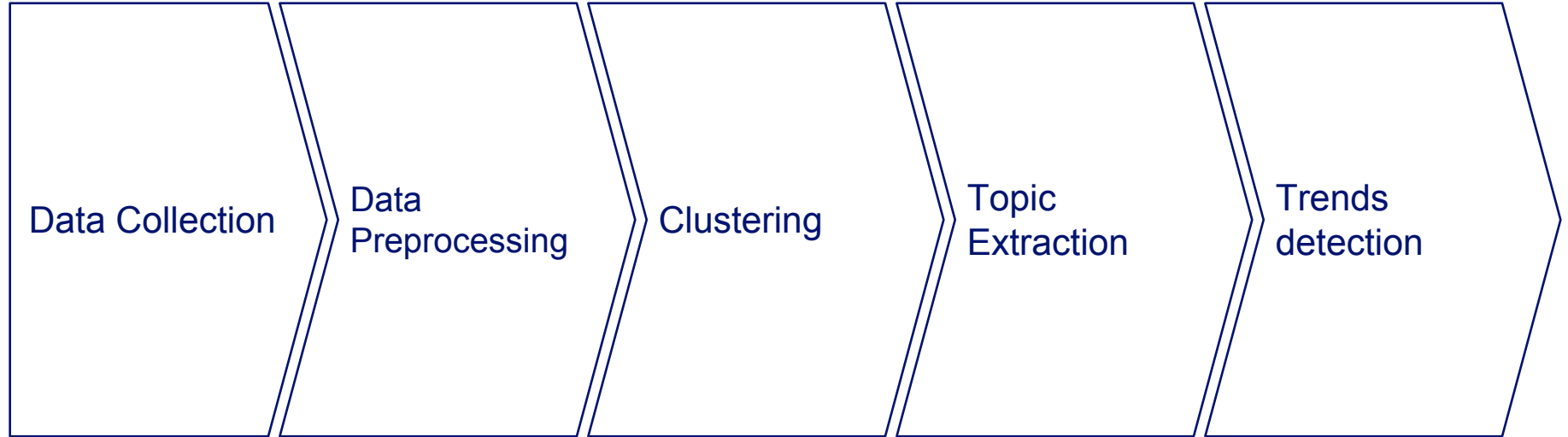
Главные тренды





ЧТО КАК СПЕЦИАЛИСТ ПО АНАЛИЗУ ДАННЫХ Я ДОЛЖЕН ДЕЛАТЬ ДАЛЬШЕ?

ПОСТРОЕНИЕ ПРОЦЕССА



СБОР ДАННЫХ

Что
популярно в
уходе за
кожей?



#beauty #skincare
#selfcare #beautyblog
 #makeup



СБОР ДАННЫХ

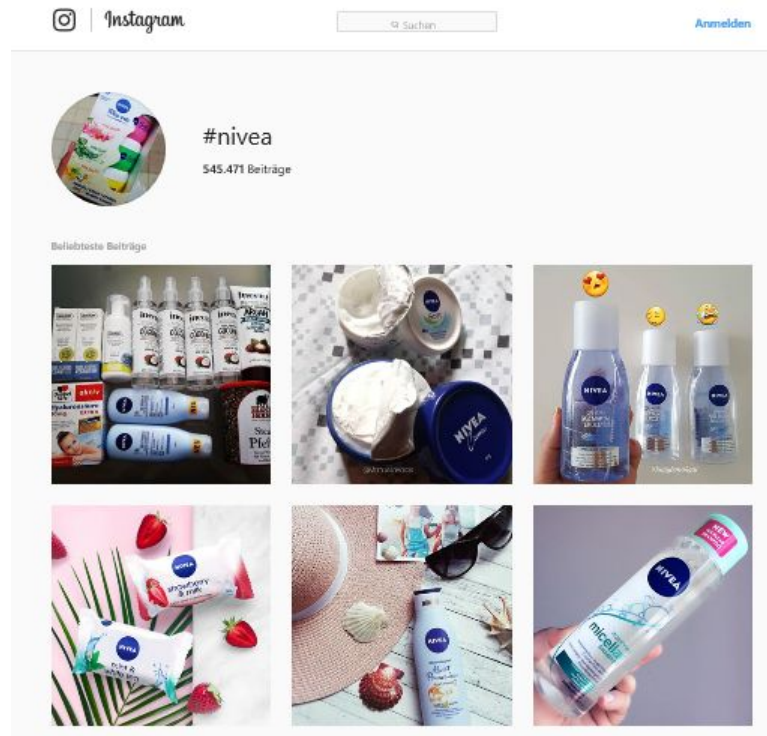
Социальная сеть	API	Ограничения
Facebook	<ul style="list-style-type: none">• Зарегистрировать приложение• Пройти долгий путь верификации	<ul style="list-style-type: none">• Facebook граф не предоставляет интерфейс поиска.• Нельзя использовать данные для сторонних исследований © Terms of privacy
Instagram	<ul style="list-style-type: none">• Зарегистрировать приложение• Пройти долгий путь верификации	НО у всех есть доступ к публичным данным
Youtube	<ul style="list-style-type: none">• Зарегистрировать приложение• Пройти не столь долгий путь верификации	Требуется переводить речь в текст

КАКОЕ РЕШЕНИЕ ПРИНИМАЕМ И ЧТО ГОВОРИМ БИЗНЕСУ?

СБОР ДАННЫХ

Социальная сеть	API	Ограничения
Facebook	<ul style="list-style-type: none">• Зарегистрировать приложение• Пройти долгий путь верификации	<ul style="list-style-type: none">• Facebook граф не предоставляет интерфейс поиска.• Нельзя использовать данные для сторонних исследований © Terms of privacy
Instagram	<ul style="list-style-type: none">• Зарегистрировать приложение• Пройти долгий путь верификации	НО у всех есть доступ к публичным данным
Youtube	<ul style="list-style-type: none">• Зарегистрировать приложение• Пройти не столь долгий путь верификации	Требуется переводить речь в текст

МЫ РЕШИЛИ ХАКНУТЬ ИНСТАГРАМ



DATA SCRAPING

Что удалось извлечь:

- дата публикации
- текст
- количество лайков
- хештеги
- немного информации об авторе: количество подписок и подписчиков

Получили: 40к постов по хештегу #scincare (2 недели)

Потраченное время: 3 дня

ТО ЧУВСТВО



КОГДА ЧЕРЕЗ ТРИ НЕДЕЛИ
НАКОНЕЦ-ТО ПОЛУЧИЛ ДАННЫЕ

ЧТО ПРЕДСТАВЛЯЕТ ИЗ СЕБЯ ТЕКСТ ПОСТА?

MENJADI
gemuk Aja
BUTUH BIAYA
MASAK PENGEN
langsingny
GRATISAN



pinkan_beautyskin • Follow

pinkan_beautyskin Happy Shopping Sista

💕💕 Cantik ngga pake sakit , Putih ,
Langsing , Dan Montok tanpa Suntik sana
sinii 🙏🏻

🍎 ORI 100%

🍎 OLSHOP AMAN

🍎 TANPA EFEK SAMPING

🍎 BERGARANSI

🍎 PRODUK TERAMPUH

🍎 DIJAMIN BERHASIL 🍎 BANYAK

PROMO DAN GIFT

Order Chat 📱 :

📞 WA Admin 1: 081907598973 (Fast
Respon)

Admin 2 : 087851073080 (Fast Respon)

#hbwhiteningori#pemutihbadan#glutamelon
e#pelangsing#olshopsurabaya#agenkosme
tik#pemutihalami#creamwajah#perawatank
ulit#lotionmurah#perawatanwajah#maskerb
adan#maskerwajah#slimbeauty#skincare#c



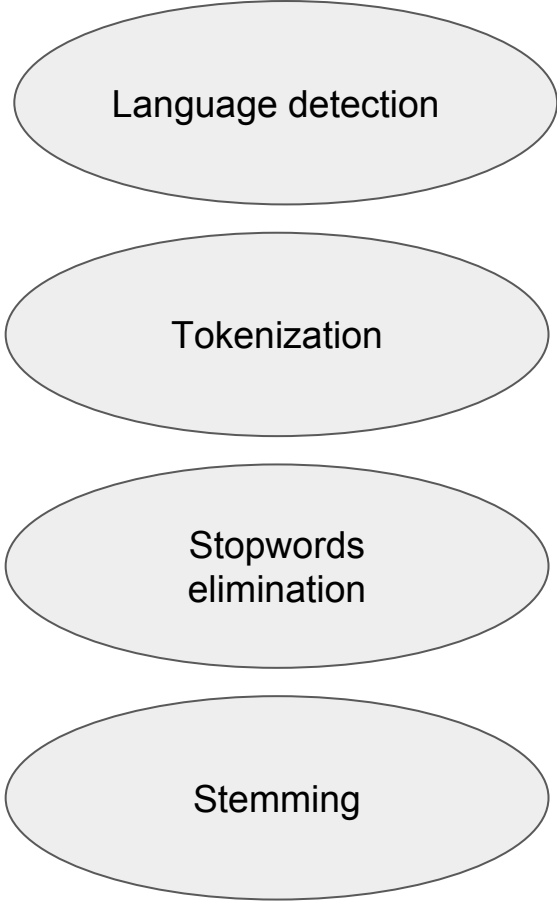
Be the first to like this.

3 MINUTES AGO

Add a comment...

...

PREPROCESSING



```
graph TD; A([Language detection]) --> B([Tokenization]); B --> C([Stopwords elimination]); C --> D([Stemming]);
```

Language detection

Tokenization

Stopwords
elimination

Stemming

ОПРЕДЕЛЕНИЕ ЯЗЫКА ТЕКСТА

- Google Translate API
- Opensource Python libraries: langdetect
- Apache OpenNLP
- свой алгоритм?

STEMMING? LEMMATIZATION?

- stemming - отсечение окончаний или типичных суффиксов

Form	Suffix	Stem
stud ies	-es	studi
stud ying	-ing	study
niñ as	-as	niñ
niñ ez	-ez	niñ

- lemmatization - приведение слова к начальной форме

Form	Morphological information	Lemma
studies	Third person, singular number, present tense of the verb study	study
studying	Gerund of the verb study	study

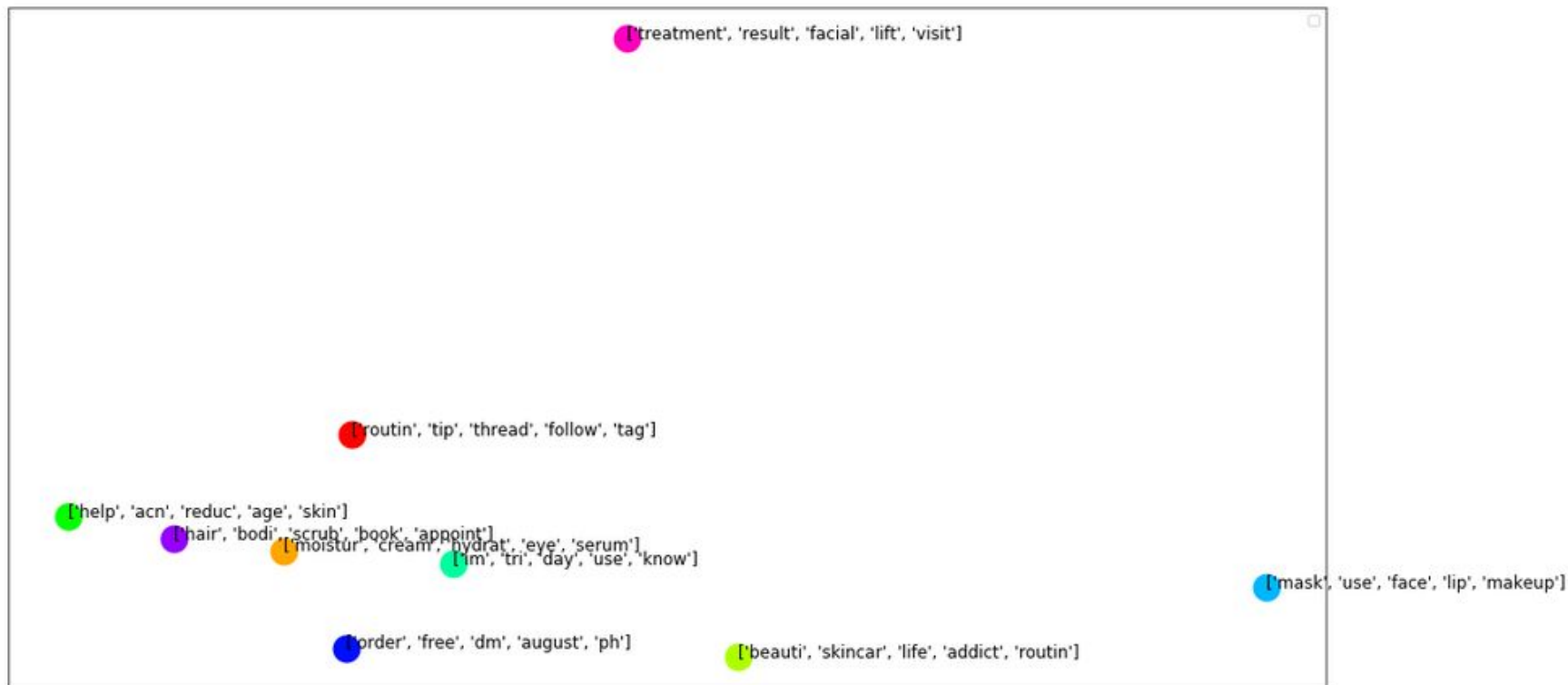
ВЫДЕЛЕНИЕ ТЕМ

Вариант “а лоб”: кластеризация + выделение популярных слов

Умный вариант : PLSA (Вероятностный латентно-семантический анализ)

BigARTM

ФИНАЛЬНЫЙ РЕЗУЛЬТАТ



ЧТО ТАКОЕ ТРЕНД?

ТРЕНД = Самый популярный?

ЧТО ТАКОЕ ТРЕНД?

«**Трендовость**» — это рост популярности некоторой темы, взвешенная в ее историческом контексте.

ЧТО ТАКОЕ ТРЕНД?

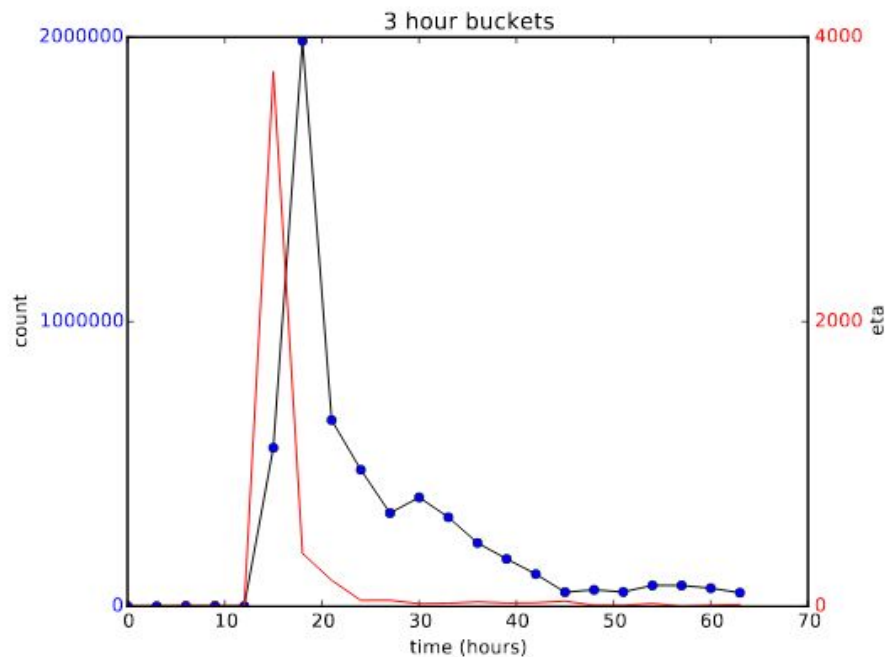
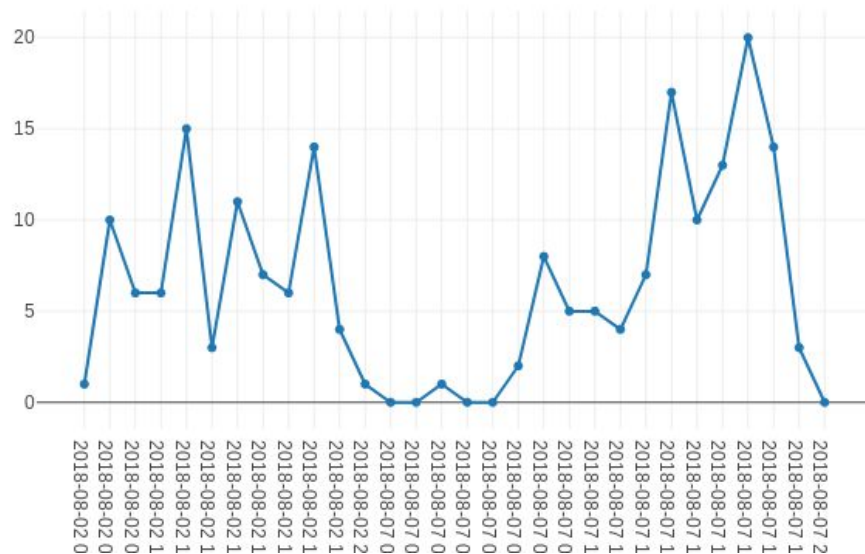


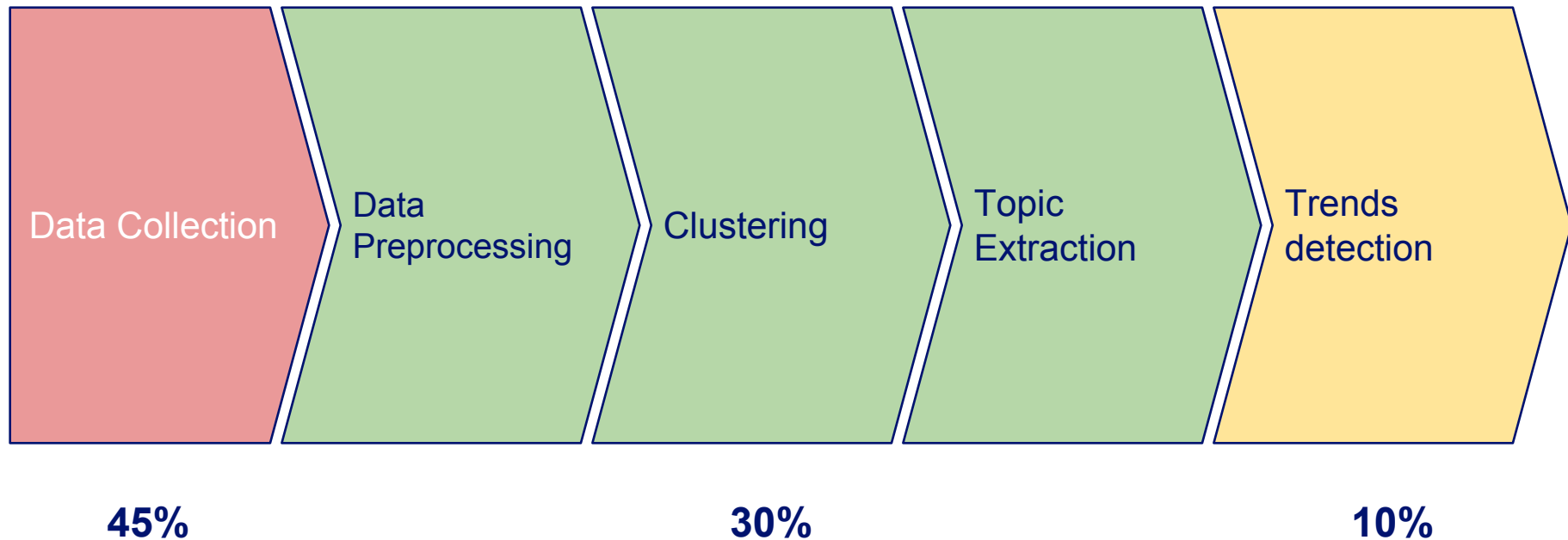
Figure 3: Mentions per 3-hour intervals of Steve Jobs, around the time of his death. The time series data are shown in the blue dots with black

СЛЕДУЮЩИЙ ШАГ



Mentions of hashtag **#treatment** per hour on 2nd July

ПРОЦЕСС: СТАТУС



СТАТЬИ

- Trends detection in social data (Twitter)
- Выделение трендов в соц сети “Одноклассники” (2017)

СПАСИБО ЗА ВНИМАНИЕ!

@dementyeva_ds