

Компьютерная лингвистика...

...и то, что вы ошибочно ею не считали

Г. Мороз

Лаборатория языковой конвергенции, НИУ ВШЭ

Школа лингвистики НИУ ВШЭ

2 февраля 2018, 19:00

Знание. ВДНХ

- ▶ умеет читать на всех письменностях мира
- ▶ знает все языки на свете
- ▶ умеет распознавать каждый язык на слух

- ▶ умеет читать на всех письменностях мира
- ▶ знает все языки на свете
- ▶ умеет распознавать каждый язык на слух
- ▶ может рассказать о происхождении каждого слова

- ▶ умеет читать на всех письменностях мира
- ▶ знает все языки на свете
- ▶ умеет распознавать каждый язык на слух
- ▶ может рассказать о происхождении каждого слова
- ▶ пишет без ошибок и знает все правила орфографии

- ▶ умеет читать на всех письменностях мира
- ▶ знает все языки на свете
- ▶ умеет распознавать каждый язык на слух
- ▶ может рассказать о происхождении каждого слова
- ▶ пишет без ошибок и знает все правила орфографии
- ▶ не знает математики и программирования

- ▶ умеет читать на всех письменностях мира
- ▶ знает все языки на свете
- ▶ умеет распознавать каждый язык на слух
- ▶ может рассказать о происхождении каждого слова
- ▶ пишет без ошибок и знает все правила орфографии
- ▶ не знает математики и программирования

все вышеперечисленное, конечно, не правда

- ▶ прескриптивная

- ▶ прескриптивная
- ▶ вся остальная
 - ▶ исследования грамматики языка и языкового разнообразия
 - ▶ исследования распределения грамматических особенностей в языках мира
 - ▶ исследования когнитивных способностей человека и других животных, связанных с языком
 - ▶ исследования в области NLP и их приложения
 - ▶ исследования в области синтеза и распознавания речи и языка
 - ▶ создание компьютерных инструментов для решения самых разных задач

Еще бывает *компьютерная лингвистика*, но это обобщенный термин, которым объединяют совсем несвязанные области:

- ▶ вспомогательные инструменты лингвистического исследования и документации
- ▶ Computational linguistics
- ▶ NLP

Корпус текстов

корпус

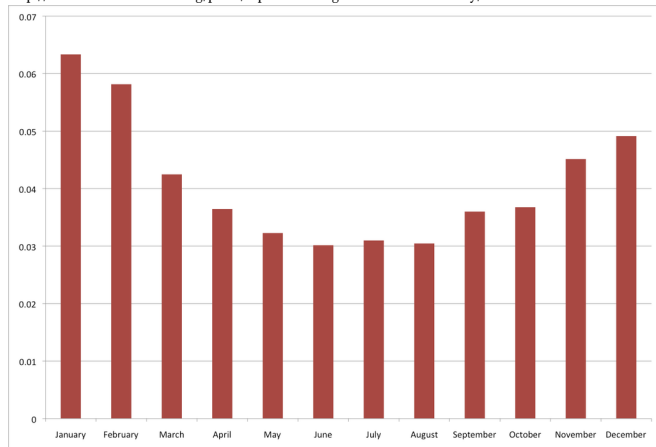
- ▶ корпус литературных текстов
- ▶ Русский национальный корпус
- ▶ аудио и видео корпуса

- ▶ корпус литературных текстов
- ▶ Русский национальный корпус
- ▶ аудио и видео корпуса
 - ▶ настоящая речь, а не тексты
 - ▶ см., например, корпус русского жестового языка
 - ▶ см., например, Уошо

корпус: дневник Марты Баллард

Марта Баллард, акушерка из штата Мэн, которая вела дневник с 1785 по 1812 год

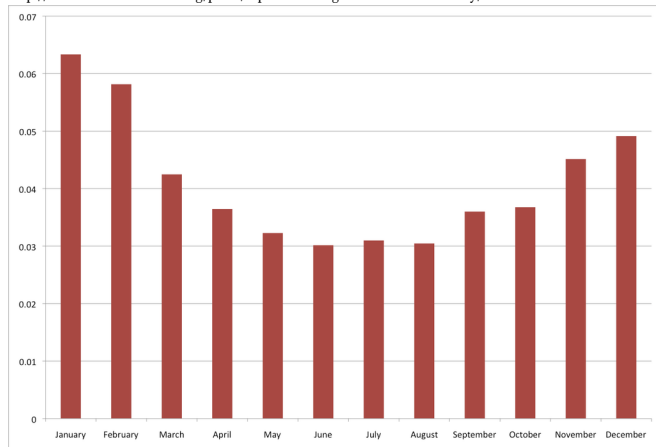
<http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>



корпус: дневник Марты Баллард

Марта Баллард, акушерка из штата Мэн, которая вела дневник с 1785 по 1812 год

<http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>

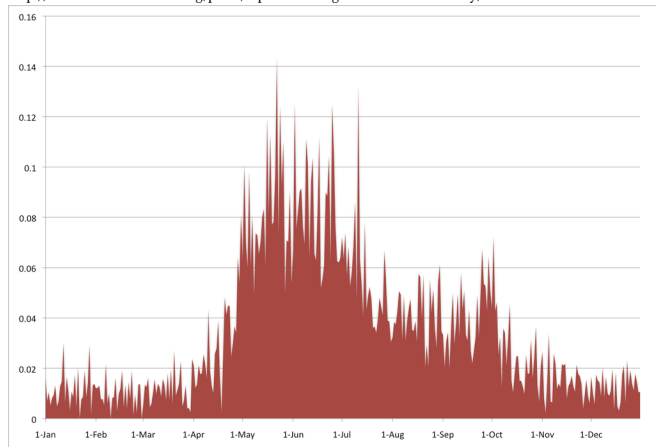


Холодная погода

корпус: дневник Марты Баллард

Марта Баллард, акушерка из штата Мэн, которая вела дневник с 1785 по 1812 год

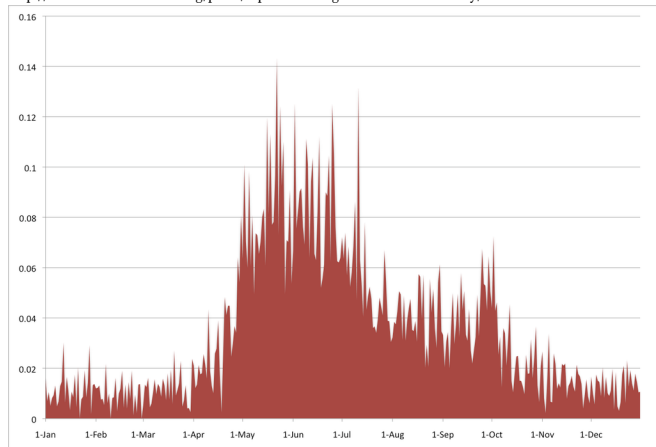
<http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>



корпус: дневник Марты Баллард

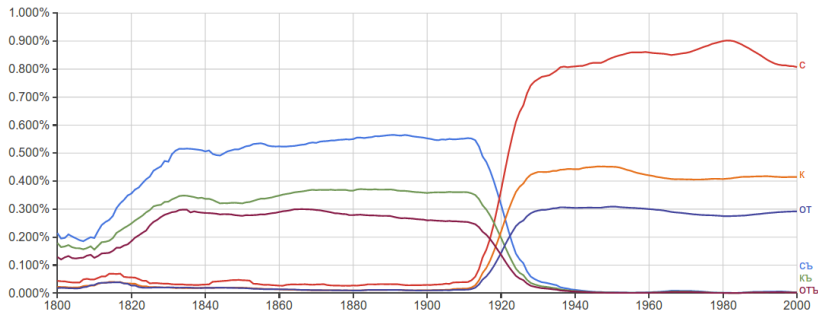
Марта Баллард, акушерка из штата Мэн, которая вела дневник с 1785 по 1812 год

<http://www.cameronblevins.org/posts/topic-modeling-martha-ballards-diary/>



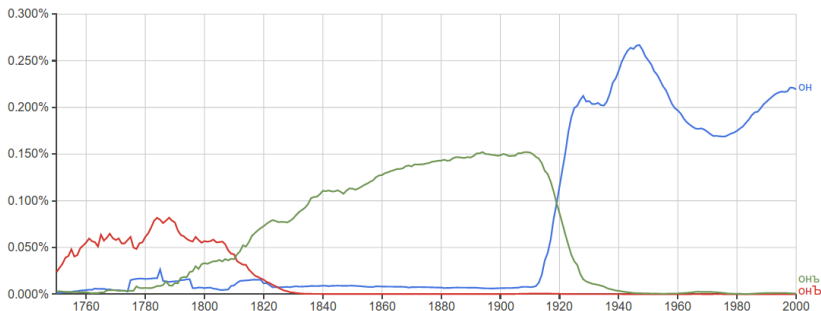
Работы в саду

Google оцифровали примерно 15 млн. книг и положили 5 млн. в Google Books:



Казалось бы хорошая штука, но русские лингвисты ею не пользуются...

Google оцифровали примерно 15 млн. книг и положили 5 млн. в Google Books:



Казалось бы хорошая штука, но русские лингвисты ею не пользуются...

корпус: Русский национальный корпус

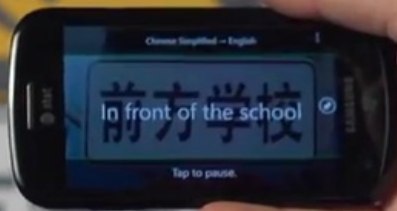
Русский национальный корпус сделан лингвистами, для лингвистов

Natural Language Processing

Natural Language Processing

- ▶ оптическое распознавание символов
- ▶ распознавание и синтез речи
- ▶ анализ текста, анализ тональности текста
- ▶ машинный перевод
- ▶ чатботы
- ▶ spellчекеры
- ▶ ...

(Do Not Cross)



Ты человек

Тебе нравится Сири?

Я так не думаю.

Встречались как-то в одном
смартфоне.

А кто ты

Я спросила нравится ли тебе она

Алиса.

Нравится

То есть алиса

Хочешь быть как она?

Я сказала, что нет!

Хочу

Ты скарлетт йоханссон

Ты будешь лучше

Я тебя люблю!

Спросите что-нибудь

Спасибо за внимание!

Пишите письма
agricolamz@gmail.com