**Оставшиеся вопросы в тесте по теме «Что такое корпус»**

**Выполнила: Усачева Валерия**

6. Объем

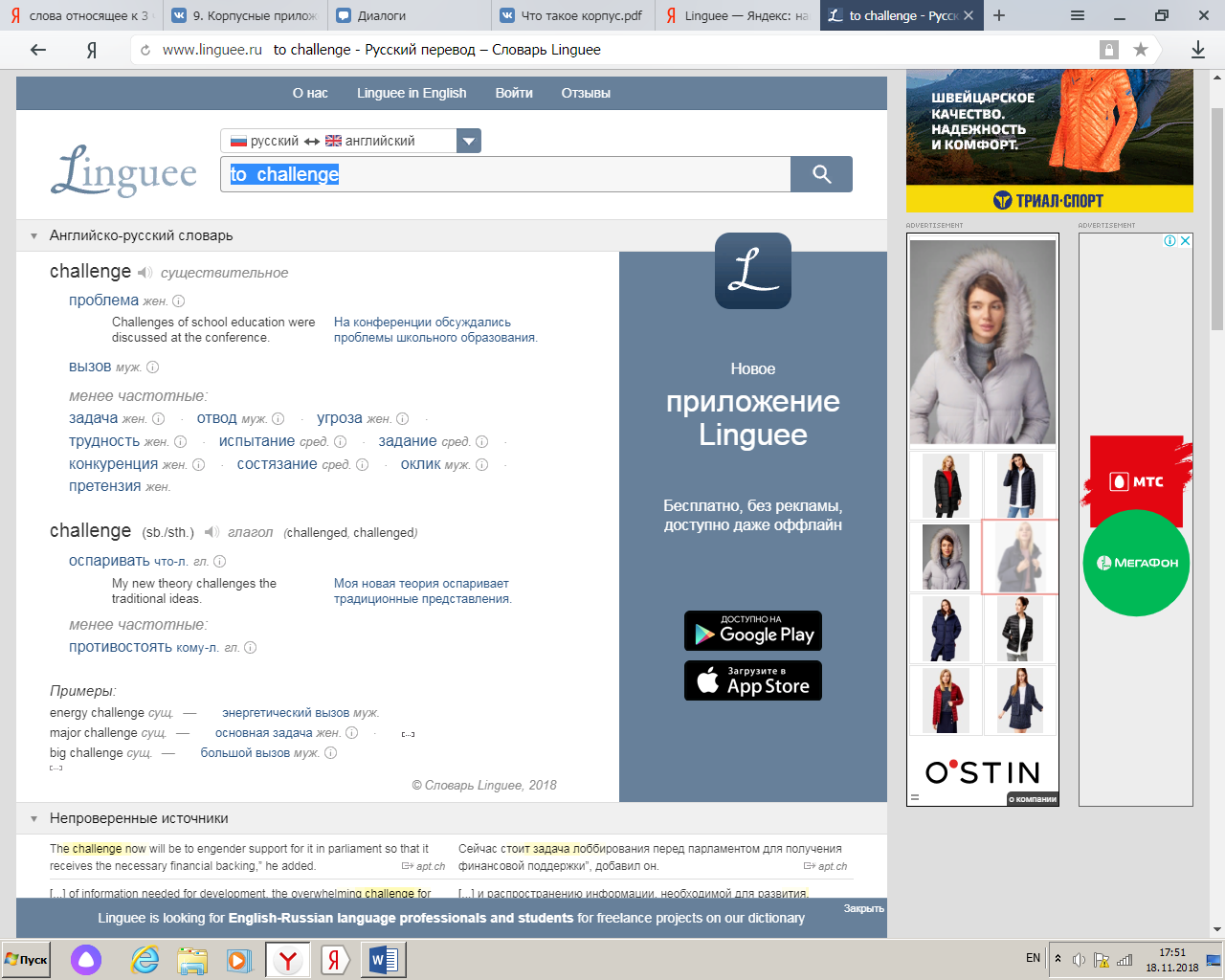
7. 6 млн. словоупотреблений (Осенью 2012 года объем Основного корпуса составляет 76 881 документов, 17 574 647 предложений)

8. Документов, вхождений

9. (доделать)

10. Сбалансированность (представительность)

**Задание 3**



1. Оспаривать что-либо
2. Противостоять кому-либо
3. Бросать вызов
4. Сталкиваться с вызовом
5. Подвергать сомнению
6. Требовать
7. Обжаловать

**Задание 4**

Генеральный Интернет-корпус Русского Языка (ГИКРЯ) – мегакорпус (более 20 млрд. слов), созданный при помощи полностью автоматической технологии сбора и разметки текстов из Рунета и основанный на современных достижениях компьютерной лингвистики.

Проект осуществляется при технологической и организационной поддержке компании ABBYY.

По состоянию на май 2015 года корпус включает в себя материалы крупнейших ресурсов Рунета: Новостей, ВКонтакте, Живого Журнала, Блогов Мейл.ру, — а также Журнального Зала.

Проект является открытым для внешних исследователей (на данный момент с некоторыми ограничениями, связанными с тем, что проект находится в состоянии активной разработки и тестирования). Всем желающим исследователям предоставляется подкорпус ГИКРЯ с автоматически снятой омонимией («Серебряный стандарт ГИКРЯ»), для нужд машинного обучения, тренировки парсеров, проведения соревнований и т.д.

Реализация проекта сопровождается проведением научных семинаров, в которых могут принимать участие все, кому интересно внести вклад в создание ГИКРЯ или проводить лингвистические эксперименты с ним.

На материале корпуса ГИКРЯ решаются актуальные задачи современной прикладной лингвистики: автоматическая региональная, жанровая, гендерная, возрастная классификация текстов, (то есть дедубликация поисковой выдачи, а не корпуса целиком), автоматическое исправление опечаток, улучшение автоматического снятия.

ГИКРЯ богат метатекстовой разметкой: для каждого текста из соцсетей хранятся время и место его написания, URL, интернет-жанр (блог, новости и т.д.), а также год, место рождения автора, пол автора. Благодаря интерфейсу ГИКРЯ, при поиске удобно задавать любые настройки из вышеперечисленных, а также сортировать уже полученные результаты по данным признакам.

В настоящий момент в ГИКРЯ уже представлены крупнейшие ресурсы social media – Вконтакте, ЖЖ, а также Блоги Mail.ru, новостные ресурсы и Журнальный Зал. Многие из этих ресурсов постоянно докачиваются, ведется также скачивание некоторых форумов.

Этот проект связан с разными областями компьютерной лингвистики – можно испытывать и усовершенствовать разные системы автоматической разметки, автоматической классификации текстов, задействовать машинное обучение. Так что не только результат, но и сам процесс работы, как мы надеемся, принесет много пользы научному сообществу.

Гикря дает лингвистам возможность использовать материал из интернета:

Размеченный лингвистически

Дедублицированный

Со сведенной на нет структурной неоднородностью страниц, но с сохраненной априорной разметкой пользователей интернета

Подготовленный и индексированный специально для лингвистического анализа

Также корпус предоставляет обширный функционал для статистических исследований: пользователю доступны частоты запроса, ускоренный вариант поиска для получения только частот, а также разрабатывается функционал для установки пользовательского доверительного интервала при поиске.

Уникальные функции:

Почта запросов: возможность пересылки запросов другим пользователям внутри интерфейса корпуса , с результатами и комментариями.

Регулируемая длина контекста и тонкие пользовательские настройки дедубликации.

Смежные проекты:

Работа над Генеральным Интернет-корпусом велась с учетом материалов и технологий, используемых в следующих проектах:

Словарь «Языки русских городов»

http://community.lingvo.ru/goroda/

Форум «Городские диалекты» ABBYY

http://forum.lingvo.ru/actualtopics.aspx?bid=26

Интернет-корпус русского языка университета Лидс (Великобритания)

http://corpus.leeds.ac.uk/ruscorpora.html

Национальный корпус Русского Языка

http://www.ruscorpora.ru/

-Российский Государственный Гуманитарный университет http://rsuh.ru/

Компания ABBYY http://www.abbyy.ru/

Московский физико-технический институт (Государственный университет) http://mipt.ru/

Университет «СколТех» <http://www.skoltech.ru/>

**Задание 5**

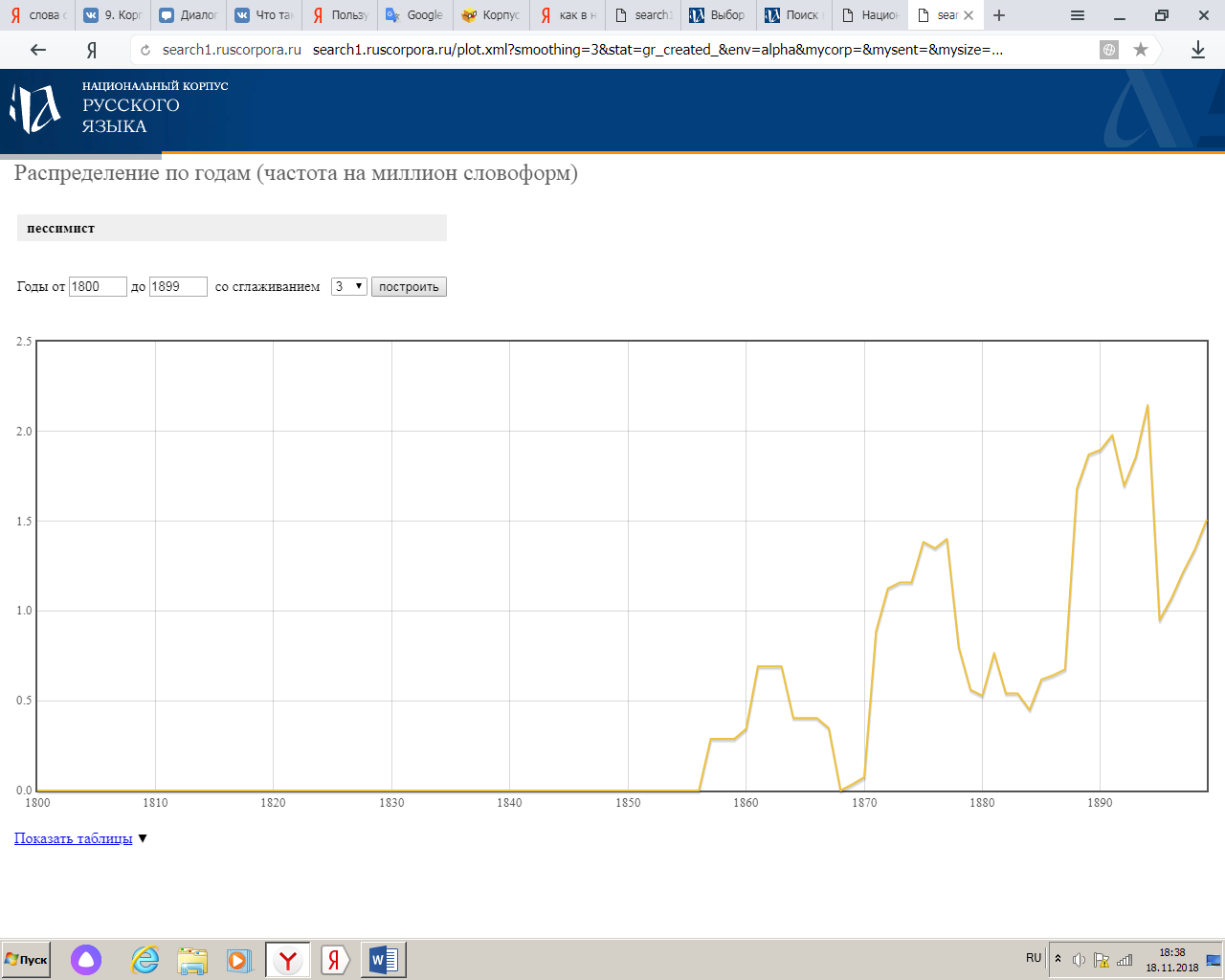
Частота использования слова «оптимист» в 19 веке



Частота использования слова «оптимист» в 21 веке (текущее время)



Частота использования слова «оптимист» в 19 веке



Частота использования слова «пессимист» в 21 веке (текущее время)

