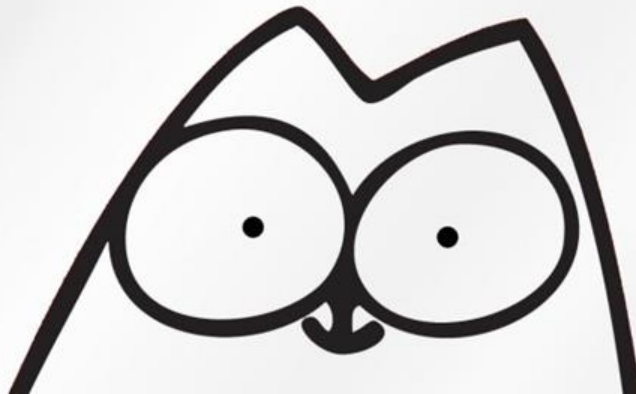


Извлечение сентиментальных слов

Руководитель: Юрий Курочкин

Выполнили: Ибрагимов Сергей, Горбунов Егор

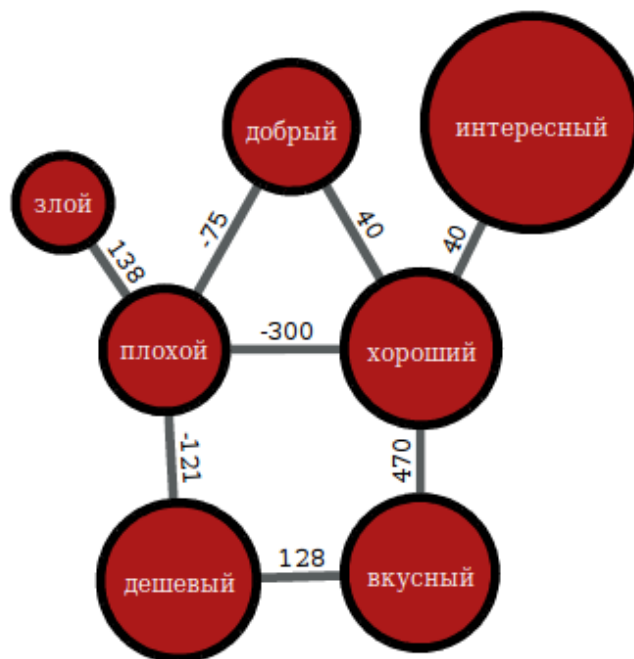


Цель

Оценить некоторые способы извлечения
сентиментальных слов применительно к
русскому языку

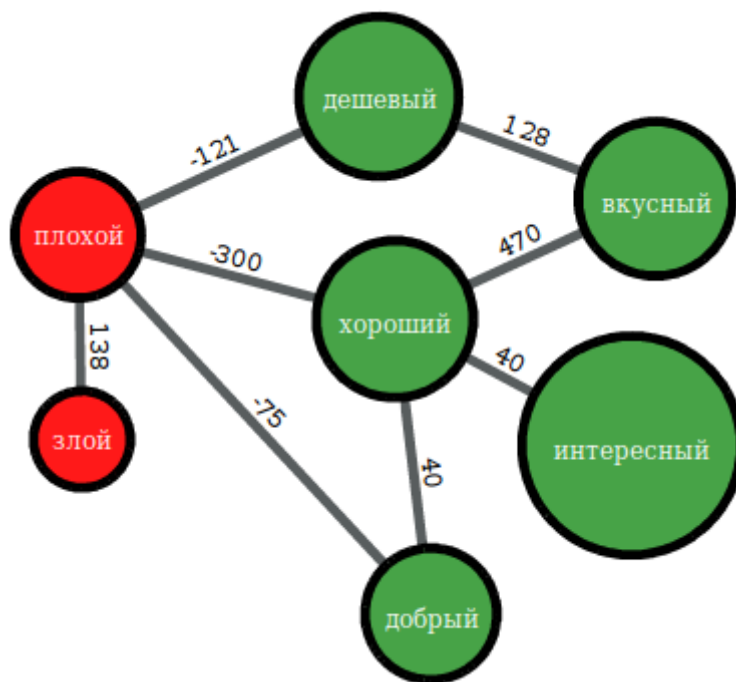
Подход

- Выделяем пары прилагательных и строим взвешенный граф



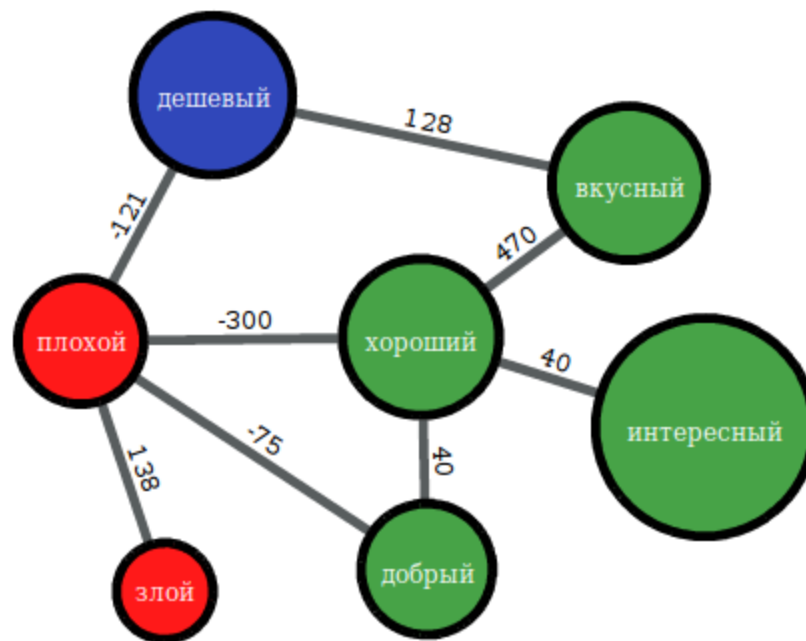
Подход

- Разделяем на 2 доли



Подход

- Выделяем нейтральные слова



Извлечение пар

Прилагательные связаны...	положительно	отрицательно
разделитель	« , » « и »	« но » « хотя » « зато »

Извлечение пар

Прилагательные связаны...	положительно	отрицательно
разделитель	«,» «и»	«но» «хотя» «зато»

Правила

- Между прилагательными...
 - расстояние < 6
 - нет глаголов и существительных
- Склеиваем «не» и прилагательное, если между ними наречия меры и степени
- Пара («хороший», «плохой») учитывается только с отрицательной связью

Извлечение пар

Прилагательные связаны...	положительно	отрицательно
разделитель	«,» «и»	«но» «хотя» «зато»

Правила

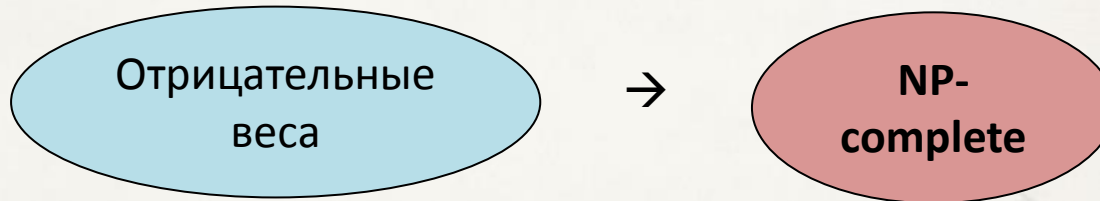
- Между прилагательными...
 - расстояние < 6
 - нет глаголов и существительных
- Склеиваем «не» и прилагательное, если между ними наречия меры и степени
- Пара («хороший», «плохой») учитывается только с отрицательной связью

Примеры

Не очень красивый, зато крепкий	→	[+1, красивый, крепкий]
Диван мягкий, вкусный сок	→	[+1, мягкий, вкусный]
Нехороший и навязчивый	→	[-1, навязчивый, хороший]

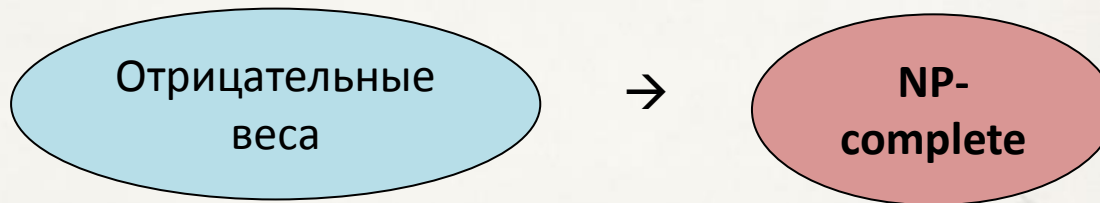
Деление графа

- Поиск минимального разреза



Деление графа

- Поиск минимального разреза

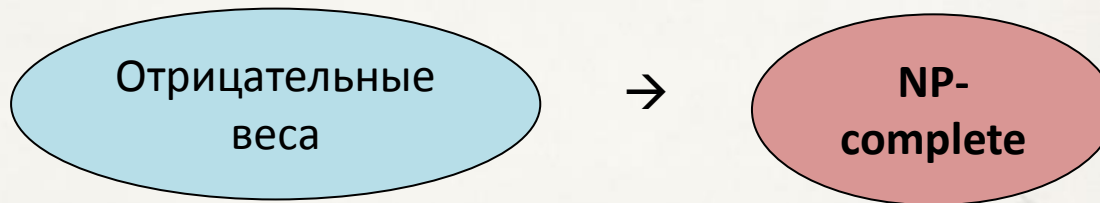


- Алгоритм Штор-Вагнер + улучшение перебрасыванием вершин

Величина разреза:	−34638
Сумма отрицательных весов:	−55000
Число «отрицательных» рёбер:	7363
Число «отрицательных» рёбер в разрезе:	5189

Деление графа

- Поиск минимального разреза

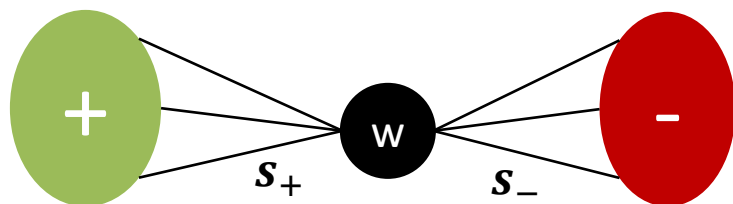


- Алгоритм Штор-Вагнер + улучшение перебрасыванием вершин

Величина разреза:	−34638
Сумма отрицательных весов:	−55000
Число «отрицательных» рёбер:	7363
Число «отрицательных» рёбер в разрезе:	5189

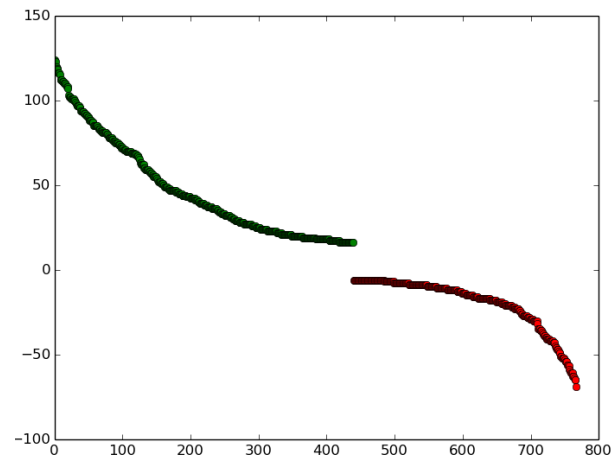
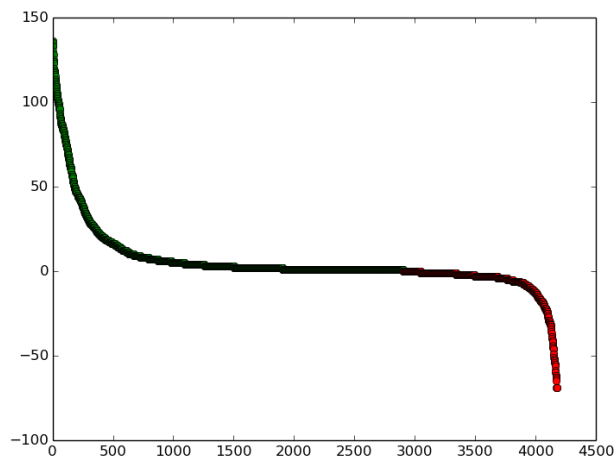
- Доля с наибольшим числом вершин - **положительная**

Выделение нейтральных слов



Притяжение вершины $w = s_+ - s_-$

- Вырезаем все вершины с притяжением в $[l, r]$ и заново пересчитываем притяжения



Нейтральных слов	3393
Положительных слов	621
Отрицательных слов	462

Результаты

Входные данные

350 мегабайт отзывов об отелях

Граф пар

4500 вершин

35000 рёбер

Размеры полученных словарей

3393 нейтральных слов

621 положительное слово

462 отрицательных слова

Результаты

Точность	62 %
Полнота	76 %

Входные данные

350 мегабайт отзывов об отелях

Граф пар

4500 вершин

35000 рёбер

Размеры полученных словарей

3393 нейтральных слов

621 положительное слово

462 отрицательных слова

Инструменты

- **C++**
- **Python**
- **mystem** - программа, которая производит морфологический анализ текста

Спасибо за внимание

