1. Титульник 8с
    a. Уважаемые члены комиссии вашему вниманию предоставляется дипломная работа на тему: автоматическое установление связей между твитами и новостными статьями.
2. Актуальность 25с
    a. Решение задачи установления связей твит-новость позволяет обогатить как твиты, так и новости.
    b. Среди преимуществ расширения новости можно выделить такие, как определение отношения аудитории к статье и получение информации для аннотирования.
    c. Многие современные методы обработки естественного языка неэффективны на коротких текстах, таких как твиты. Собственно, главным преимуществом расширения твита с помощью новости является возможность использования большего количества методов обработки естественного языка, среди них тематическое моделирование, классификация, выявление тональности.
3. Постановка 24с
    a. Задачей данной работы является автоматическое установления связей между твитами и новостными статьями.
    b. Это включает в себя сбор и разметку данных, написание программного обеспечения, позволяющего установить связи, исследование с целью нахождения наилучшего подхода.
    c. Ввиду особенностей используемых данных перейдём к понятиям применимым в информационном поиске. Будем рассматривать твит как запрос, а список новостей как ответ. То есть для каждого твита мы получаем список новостей, упорядоченный по мере убывания схожести. Будем называть такой список рекомендацией,
4. Получение данных 27с
    a. Для начала необходимо собрать данные: твиты и новости.
    b. Сбор твитов осуществляется через Twitter Streaming API - открытый API, предоставляемый твиттером. Это позволяет получить 1% от всех публикуемых твитов.
    c. Сбор новостей осуществляется через RSS-каналы. В рамках работы были собраны новости из 5 наиболее популярных в твиттере новостных источников. Но всё равно в собранных твитах только 3% ссылок введут на выбранные источники.
    d. В итоге получено множество состоящее из полумилиона твитов и 14 тысяч новостей.
5. Набор данных
    a. Есть автоматическая, заключается в следующем на рисунке тото
    b. Нетривиальные такие-то
    c. Получено только 746 нетривиальных связей, поэтому ручная
    d. Ручная описание рисунок
6. Метод TF-IDF
    a. Для построения рекомендаций на основе размеченного набора данных в работе используется три метода.

   b. Все используемые методы построения рекомендаций для каждого текста порождают вектор, с помощью которого можно определить схожесть двух текстов.
   c. Собственно, метод TFIDF для каждого слова в каждом из тестов расчитывает значение tfidf и строит следующую матрицу. В матрице строки соответствуют - словам, а столбцы - текстам, в ячейке записано значение tfidf.

7. Метод WTMF
8. Метод WTMF-G
9. Построение моделей и оценка качества
10. Построение рекомендаций
11. Используемые в работе метрики
12. Результаты экспериментов
13. Пример результата
14. Заключение