Извлечение аспектов и Twitter Классификация настроений по правилам фрагментов

Васильев В.Г. (vvg_2000@mail.ru),
Денисенко А.А. (denisenko_alec@mail.ru),
Соловьев Д.А. (dmitry_soloviev@bk.ru)

ООО «ЛАН-ПРОЕКТ» , Москва, Россия

В статье рассматриваются подходы к извлечению явных аспектов из отзывов пользователей о ресторанах и классификации настроений в сообщениях Twitter телекоммуникационных компаний на основе правил фрагментов. В этой статье представлена модель правила фрагмента для классификации настроений и явного извлечения аспектов. Правила могут создаваться экспертами вручную и автоматически с использованием процедур машинного обучения. Мы предлагаем алгоритм машинного обучения для классификации настроений, который использует термины, созданные правилами фрагментов, и некоторые методы, основанные на правилах, для явного извлечения аспектов, включая метод, основанный на генерации правил фильтрации. В статье представлены результаты экспериментов на тестовом наборе для классификации настроений телекоммуникационных компаний в Твиттере и извлечения явных аспектов из отзывов пользователей о ресторане. В документе сравниваются предложенные алгоритмы с базовыми и лучший алгоритм для отслеживания. Обучающие наборы, метрики оценки и эксперименты используются в соответствии с SentiRuEval. В качестве нашей будущей работы можно выделить такие направления, как: применение полууправляемых методов генерации правил для снижения трудозатрат, использование активных методов обучения, построение системы визуализации генерации правил, способной обеспечить процесс взаимодействия с экспертами.

Ключевые слова: правила фрагментов, классификация настроений, извлечение аспектов, анализ мнений.

1. Введение

Интеллектуальный анализ мнений и извлечение настроений — активно развивающаяся поддисциплина интеллектуального анализа данных и компьютерной лингвистики. Перспективный подход к автоматическому извлечению настроений основан на выделении конкретных характеристик продукта — аспектов и на определении этих полярностей. Обычно проблема решается в три этапа. Сначала извлекаются аспекты и те полярности. Затем аспекты переключаются на категории, если они предопределены. В противном случае набор аспектов группируется и выбираются репрезентативные аспекты. Заключительный этап включает классификацию полярности категорий на основе полярности отдельных аспектов.

В этой статье мы представляем подход, основанный на правилах, который использует модель правил фрагментов для явного извлечения аспектов из обзоров пользователей и для классификации настроений сообщений Twitter. Основным преимуществом подхода является его хорошая интерпретируемость.

С одной стороны, есть возможность использовать экспертные знания в модели за счет ручного построения правил. С другой стороны, вы можете построить модель автоматически или получить интерпретируемую модель в рамках процедуры, включающей взаимодействие эксперта и системы.

В работе [7] описаны подходы к классификации настроений рецензий на фильмы. Эти подходы основаны на подсчете количества предлагаемых положительных и отрицательных слов и использовании наивного байесовского классификатора, классификации с максимальной энтропией, метода опорных векторов. Использование метода опорных векторов повышает точность до 82%.
Еще два метода классификации дают точность 75–80%. В статье [1] описана классификация настроений в Твиттере на основе метода опорных векторов. В качестве признаков используются слова, словосочетания и части речи. Результаты, показанные в этой статье, совпадают с результатами, показанными в предыдущей статье, и подчеркивается, что использование части речи не повышает точность.

В статье [2] рассмотрены два подхода к классификации настроений кинорецензии. Первый подход основан на количестве положительных и отрицательных терминов, интенсификации терминов и изменении семантической полярности конкретного термина. Второй подход использует алгоритм машинного обучения, машины опорных векторов. Использование первого подхода дает точность около 65–70%. Использование второго подхода повышает точность до 85%. Сочетание двух подходов не увеличивает точность.

В статье [3] авторы предлагают подход к классификации настроений с обнаружением смещения полярности. Предложения со смещенной и несмещенной полярностью используются в качестве признаков для классификации на основе метода опорных векторов. Такой подход позволяет несколько улучшить качество по сравнению с базовым.

В дополнение к словарному и векторному подходу к классификации настроений в ряде работ предлагаются специальные вероятностные модели, например, древовидная классификация настроений и использование отношений между словами [6]. Также в ряде работ авторы четко определяют правила оценивания текстов. В частности, в статье [7] сформулированы различные правила определения области применения обратных слов, таких как «нет» . Так, в работе по классификации тональностей используются как стандартные методы классификации текстов, так и модифицированные методы, учитывающие полярность сдвинутых терминов, синтаксическую структуру предложений, отношения между словами.

В настоящей статье подход к классификации настроений в Твиттере основан на функциях, извлеченных с использованием правил фрагментов. Полученные таким образом признаки при правильном задании правил образуют пространство меньшей размерности и обладают хорошей описательной силой, как это было показано в [10].

Анализ мнений на основе аспектов широко исследовался. Известны несколько подходов к решению этой задачи [4]: (1) частотный подход, (2) подход, основанный на правилах, (3) методы обучения с учителем, (4) методы тематического моделирования.

Частотный подход использует тот факт, что 60–70% аспектов являются явными существительными [4]. Утверждается, что люди пишут обзоры на аспектном языке, потому что они также читают другие обзоры и принимают терминологию. Подход, основанный на правилах, использует предположение о том, что существует какая-то связь между аспектами и полярностями, выраженными в тексте. Отношение можно формализовать с помощью правил. Существует также гибридный подход, выражающийся в использовании правил фильтрации извлеченных именных словосочетаний.

Проблема может рассматриваться как проблема маркировки последовательностей в соответствии с некоторыми предлагаемыми методами машинного обучения с учителем. В частности, Скрытая Марковская Модель

можно использовать условные случайные поля. Методы тематического моделирования используют естественное предположение, что темы обзоров являются соответствующими аспектами.

В этой статье предлагается основанный на правилах подход к извлечению аспектов. Существуют две основные модели правил: основанные на грамматике и на основе фрагментов. Модели грамматики включают в себя применение контекстно-свободных грамматик, например парсер Tomita [8]. Другая модель основана на использовании специальных фрагментов текста и представляет собой ряд операций над этими фрагментами. Правилом в этом случае является декларативное описание извлекаемой информации. Наша модель является примером последнего подхода.

В связи с тем, что отзыв извлечения аспектов может быть достигнут с помощью различных словарей, таких как тезаурус и словари для предметной области, важным вопросом является повышение точности. В данном случае улучшения выражаются в использовании специальных механизмов фильтрации извлеченных аспектов. Здесь можно использовать, в частности, правила фрагментов. Целью участия в треке было тестирование основанных на правилах фрагментов подходов к извлечению аспектов и классификации твитов. Кроме того, мы попытались использовать методы автоматической генерации фрагментных правил.

Остальная часть статьи выглядит следующим образом. В разделе 2 дано формальное описание языка правил фрагментов и описание предлагаемых подходов. В разделе 3 анализируются полученные результаты; дано сравнение с исходными результатами и лучшими результатами трека. В разделе 4 представлены выводы и дальнейшая работа.

2. Методы

2.1. Модель правил фрагмента

В данной работе для описания признаков текста и правил классификации мы использовали математический эматическая модель, основанная на определении операций над множествами текстовых фрагментов [9].

Пусть у нас есть текст = (1, ...,), где — единичный элемент текста, = { 1, ..., } — множество всех элементов, — длина текста, — количество различий
отдельные элементы текста.

Определение

1 Множество = { (,) | 1 } будем называть множеством всех частей длины текста. Фрагментами текста будем называть единичные элементы множества = (,) , задающие левый и правый граничные фрагменты (номера первого и последнего элементов во фрагменте).

Определение

2. Пусть = (,) и = (,) , тогда | | = + 1 — длина фрагмента; если и — отношение , включения; if < or = & < — отношение порядка.

,

Определение 3

Множество фрагментов будем называть редуцированным, если таких , что .

( ) обозначают редуцированное множество фрагментов на основе множества , — сократить операцию.

Определение 4

Расстояние между фрагментами = (,) и = (,) определяется следующим образом:

Определение 5

Результатом правила a для текста является множество , содержащий все фраг мент, относящийся к этому правилу. Если , то назовите текст релевантным правилом .

Определение 6

Основные правила это правило = , , результатом которого является = { 1, ..., } — редуцированное фрагментов, элементов, выделяющихся за одну операцию. Комплексное правило — это правило , которое получается выполнением операций над другими

Определим теперь возможные операции для построения комплексных основные правила 1, ..., .

Определение

7 = 1 2 — бинарная операция ИЛИ,

Например, правило извлечения фрагментов хорошего качества наилучшего качества имеет отношение к появлению количество этих слов в тексте.

Определение

8 = 1 — бинарная операция И с ограничением на расстояние между фрагментами,

Например, правило извлечения фрагментов хорошего качества наилучшего качества имеет отношение к появлению слов, эту операцию можно использовать без каких-либо менее 4 ограничений на расстояние между словами.

Определение

= 1 2 — бинарная операция последовательности с ограничением на расстояние между = осколков, расстояние

Например, правило извлечения фрагментов хорошего качества наилучшего качества имеет отношение к появлению «распродажа» или «рас (1, 2) > 0, 1 1 , 2 счет» . Эту операцию можно без каких-либо ограничений на расстояние между словами. (1, 2) 2}

Извлечение последовательностей по правилам фрагментов

Определение

кратные последовательности (последовательности фрагментов)

(например, правило ... «сотовый мегафон»), соответствующих разным руководителям связи.

Определение

11 — бинарная операция нахождения пересечения фрагментов,

Например, правило [Chapter $SentBegin] извлекает слова “Chapter”, которые пишется в начале предложения.

Определение 12

— унарный оператор накладывает ограничения на длину фрагмента,

Например, правило ... #INTEGER ...

специальные варианты бинарных правил, в котором один из операндов считается отрицательным или условным. Например, является оператором, находящим последовательность, в которой второй операнд берется из отрицания; является оператором, находящим последовательность, в которой первый операнд берется из отрицания; — оператор найти последовательность, в которой первый операнд является условным.

Правило определяется как

Например, правило № ^:3 (хорошее лучшее качество) извлекает слова «хороший», «лучший» и «качество», перед которым нет слова «нет» на расстоянии трех слов.

#define задает именованное выражение в тексте правил ... нажатие подставляется в текст правила. Эти выражения используются, чтобы избежать повторения элементов сложных правил ... сохранение результатов поиска ... обращении к переменной производится ... использование именованных выражений ... в правиле необходимо использовать операторы @ и @@.

Например, #define Good (хорошее лучшее качество) задает именованное выражение Good, которое следует обрабатывать @Good.

2.2. Классификация настроений

Для классификации настроений мы использовали гибридный подход, основанный на сочетании извлечения признаков на основе правил и обучения классификатора методами машинного обучения. Индукция классификатора включает в себя предварительную обработку обучающего набора, извлечение признаков с использованием предопределенного набора правил фрагментов, обучение классификатора с использованием выбранных методов машинного обучения.

Тексты в обучающей выборке предварительно обрабатываются с использованием следующих процедур: 1. Графематический анализ (токенизация, определение границ предложений, фонетическое кодирование, извлечение дескрипторов слов).

2. Лингвистический анализ (лемматизация, частеречевое тегирование, снятие неоднозначности смысла слов, извлечение словосочетаний, извлечение синтаксических признаков).

3. Построение индексов нижнего уровня (инвертированный индекс исходных словоформ, инвертированный индекс индекс леммных словоформ, инвертированный индекс дескрипторов слов).

Общая схема алгоритма обучения имеет следующий вид.

1. Построение векторного представления текстов с помощью набора правил фрагментов.
2. Уменьшение размеров и расчет весов функций.
3. Обучение и оценка классификатора на обучающей выборке.

На первом этапе для извлечения признаков используется предопределенный набор из 100 специальных правил фрагментов.

Пример правила фрагмента:

@@COND^:5((@@NEG^:5\s(@@INTENS^:5\s($Adj $Verb $Noun $Adv))) &5\s? @@OBJECT) ,

где @@COND — слова-условия («если»), @@NEG — отрицательные слова, @@INTENS — в напряженных словах («очень» , «далеко» , «чисто»), @@OBJECT — объект («мц» , «мегафон» , «билайн»).

На втором этапе мы использовали общепринятые методы уменьшения размерности и вычисления весов признаков.

На третьем этапе обучаются два классификатора, один классификатор для положительного класса и один для отрицательного класса. Для обучения классификатора мы использовали нашу надежную реализацию следующих стандартных методов машинного обучения:

1. Байесовский классификатор на основе многомерного распределения Гаусса (gmm), 2.
Классификатор K-ближайших соседей (knn), 3.
Классификатор Мизеса-Фишера (vmfs), 4.
Классификатор Роччио (roccio), 5.
Классификатор опорных векторов (свм).

Обученные положительные и отрицательные классификаторы используются для построения окончательного решающего правила следующего вида:

$$\begin{matrix} ( ) > ( ) \mid ( ) < ( ) & ( ) = ( ) = 1, ( ) = & ( ) > ( ) ( ) < \\ 1, \quad \mid ( ) = ( ) = 0 & ( ) = 1, & ( ) \end{matrix}$$

$$' ( ) = \text{\textcircled{1}},$$

где ′(·) { 1, 0, 1} — окончательное решающее правило, — ( ) {0, 1} и ( ) ( ) {0, 1} ( ) решающие правила для положительного и отрицательного классов, [0, 1] и [0, 1]

степень соответствия положительному или отрицательному классу (для вероятностных классификаторов — вероятность отнесения к соответствующему классу, для svm — расстояние до соответствующей гиперплоскости и т. д.), — набор признаков в тексте.

2.3. Явное извлечение аспектов на основе правил

Существует два типа аспектов, определенных в анализе мнений на основе аспектов: явные и неявные. Явные аспекты - это понятия, которые явно упоминаются в предложении. Неявные аспекты выражаются косвенно. В этом разделе предлагается ряд подходов к явному извлечению аспектов на основе правил фрагментов. Предварительно пусть = { 1, ..., } — множество уникальных аспектов, извлеченных экспертами и представленных в обучающей выборке.
Тренировочный комплект предоставлен организаторами SentiRuEval [5].

Множественная операция ИЛИ

В основном, для явного извлечения аспекта можно использовать такое правило фрагмента:

$$= \quad (1, 2, ...,), \quad .$$

Здесь — правило, где операция ИЛИ выступает связующим звеном между уникальными аспектами. Фактически для каждого аспекта извлекается соответствующий набор фрагментов. Результатом операции является редуцированный единый набор фрагментов.

Многократное ИЛИ с максимальным уменьшением В данном случае

может возникнуть следующая ситуация. Вместо целого в целом могут быть извлечены структурные части. Например, есть три выделенных аспекта ГОРЯЧЕЕ, БЛЮДО, ГОРЯЧЕЕ БЛЮДО. Стандартный метод редукции удалит самый большой фрагмент ГОРЯЧЕЕ БЛЮДО, и мы получим два аспекта вместо одного. В связи с этим было принято решение модифицировать метод редукции и исключить фрагменты, входящие в состав других фрагментов. Также следует отметить, что соседние фрагменты могут быть одним аспектом.

Поэтому перекрывающиеся фрагменты и соседние фрагменты должны быть объединены.
В результате извлекаются фрагменты максимальной длины.

Фильтрация на основе

правил Также представляется уместным использовать фильтрацию на основе правил для извлечения аспектов. Алгоритм извлечения построен следующим образом. Сначала с помощью выбранных экспертами аспектов извлекаются фрагменты от аспекта до ближайшего прилагательного. Затем на основе извлеченных фрагментов (шаблонов) формируются наиболее общие правила. Здесь в пространстве признаков определено ранее. Сгенерированные правила применяются для фильтрации набора извлеченных аспектов-кандидатов путем подсчета поддержки и удаления кандидатов с поддержкой ниже порогового значения. Как уже упоминалось, припоминание может быть достигнуто с помощью соответствующих словарей. В этом случае процесс фильтрации необходим для повышения точности. Определение контекста некоторых аспектов позволяет отделить ситуации, когда термин не является аспектом.

Пусть ( ) — правило, результатом является набор фрагментов от вида до ближайшего прилагательного. Алгоритм извлечения аспекта для каждого аспекта, выбранного экспертами, генерирует набор контекстов аспекта ( ) путем применения правила ( ) к обучающему набору.

Затем алгоритм генерации правил строит шаблоны этих контекстов. В каждом обзоре аспекты-кандидаты извлекаются и фильтруются с использованием этих шаблонов. Наконец, у нас есть набор извлеченных явных аспектов.

Алгоритм2. Явное извлечение аспектов с фильтрацией — набор данные. аспектов, выбранных экспертами Входные — иерархия признаков, — набор поездов, — тестовый набор

Выход. — извлечены явные аспекты.

Шаг 1. Для всех (, ( ))

Шаг 2. Для всех
Для всех
(( ))

Существует ряд классических алгоритмов поиска наборов часто встречающихся элементов, которые используются для генерации правил, таких как Apriori, FP-growth, Eclat. Важным отличием этих алгоритмов является способ представления данных. В основном есть два подхода — горизонтальное и вертикальное представление. В вертикальном представлении необходимо иметь списки фрагментов, соответствующих элементам правила. В горизонтальном представлении каждому фрагменту соответствует набор элементов правила. Вертикальное представление более практично в случае фрагментарной модели. В этом контексте можно применить один из известных алгоритмов — Eclat [11]. Тем более, что поддержка правил определяется пересечением множеств фрагментов.

Для фильтрации используются правила вида 1 1,1 2 1,1... 1,1 . Поиск правил основан на иерархии признаков. В качестве элементов иерархии могут быть дескрипторы частей речи, отдельные слова и т.п. Последовательно от дескриптора $Any (любое слово) разворачивается и уточняется правило. Критерием выбора является степень специфичности правил и минимальный порог поддержки. Специфика правил возрастает в зависимости от количества элементов и их места в иерархии. Чем больше элементов и чем ниже место элементов в иерархии, тем выше специфичность.
В этом случае правила исключаются при поддержке ниже порога. В результате каждый аспект связан с набором правил. Таким образом, фильтрация осуществляется, когда есть только те аспекты-кандидаты, которые соответствуют хотя бы одному правилу.

3. Оценка

3.1. Классификация настроений в Твиттере

Для обучения используется обучающий набор, состоящий из 3846 твитов телекоммуникационных компаний. Каждая компания, упомянутая в Твиттере, оценивалась по шкале { 1, 0, 1}.

Тестовый набор состоит из 5322 твитов о телекоммуникационных компаниях. Целью тестирования было отнести каждое упоминание компании к одному из трех классов: положительному, отрицательному или нейтральному. Показатели макро-мера и микро-мера используются для оценки качества. Результаты тестирования представлены в Таблице 1. В таблице показан лучший метод, базовый уровень и 5 прогонов:

9_1 Байесовский классификатор, основанный на сочетании многомерных нормальных распределений (gmm),
классификатор 9_2 k-ближайших соседей (knn), 9_3
байесовский классификатор на основе распределения фон Мизеса-Фишера (vmfs), 9_4
центроидный классификатор Roccio (roccio),
9_5 классификатор на основе машин опорных векторов (svm).

Базовый уровень относит все твиты к наиболее частому классу, в данном случае к отрицательному. Для обучения используется обучающая выборка, состоящая из 3846 твитов телекоммуникационных компаний. Каждая компания, упомянутая в Твиттере, оценивалась по шкале { 1, 0, 1}.

Показатели макро-мера и микро-мера используются для оценки качества [5].

Таблица 1. Оценка качества твитов с классификацией настроений

Алгоритм 9_1	Макро-мера	Микро-мера
(gmm) 9_2	0,3158	0,3331
(knn) 9_3	0,2328	0,2626
(vmfs) 9_4	0,3305	0,3371
(roccio) 9_5	0,3310	0,3501
(svm)	0,3527	0,3765
Базовый уровень	0,1823	0,3370
2_Б	0,4829	0,5362

Оценка качества классификации осуществляется на уровне базового микропоказателя и значительно выше макропоказателя. Это можно объяснить особенностью базовой линии и правилом расчета микро- и макромер. Макро-мера — это среднее количество нормативно-меры, рассчитанное отдельно по трем классам. Базовый алгоритм имеет нулевую меру для двух классов (положительный и нейтральный), но отрицательный класс -меры имеет значение около 55%. При усреднении трех классов мера оказывается равной 18%. Наш алгоритм решает эти проблемы. Алгоритм, основанный на методе опорных векторов, показал лучшее качество. Алгоритм на основе k ближайших соседей показал худший результат. Как видим наш результат сравним с результатом других участников.

3.2. Явное извлечение аспекта

Оценка результатов проводилась по тренировочному набору (золотой стандарт), предоставленному организаторами. Комплект состоит из 202 аннотированных рецензий на русском языке. Мы использовали стандартные меры: точность, полноту и F-меру. В официальных результатах метод, основанный на многократном ИЛИ с максимизирующим сокращением, имеет идентификатор — 11.1.

Таблица 2. Результаты оценки явного извлечения аспектов

Метод	Строгие требования			Слабые требования		
	п	р	F1	п	р	F1
ИЛИ	49%	71%	58%	59%	72%	65%
Многократное ИЛИ с максимальным уменьшением [11.1]	66%					
Фильтрация на основе правил	60%	64%	62%	66%	69%	67%
Базовый уровень	55%	69%	61%	65%	70%	67%
[2.1] Лучший результат/сильный	69%	55%	69%	61%	69%	79%
[4.1] Лучший результат/слабый						

Метод	Строгие требования			Слабые требования		
	п	р	F1	п	р	F1
ИЛИ	51%	73%	60%	61%	74%	72%
Многократное ИЛИ с максимальным уменьшением [11.1]						
Фильтрация на основе правил						
Базовый уровень	72%	57%	63%	81%	62%	71%
[2.1] Лучший результат/сильный	73%					
[4.1] Лучший результат/слабый						

В целом участники официального трека имели сопоставимые результаты. Получается, что подход, основанный на переносе аспектов из набора поездов в набор тестов с нормализацией, показывает те же результаты, что и подходы, использующие сложные модели для обучения.

Результаты показывают, что модификация операции множественного ИЛИ обычно способствует повышению производительности. Можно утверждать, что максимизация редукции показала преимущество по сравнению с минимизацией редукции, когда есть только те фрагменты, которые не содержат других. Это сокращение применяется при решении задач классификации текстов и дает преимущества с точки зрения скорости выполнения правил классификации. В будущем различные типы сокращения могут принимать форму отдельных операций вместо использования по умолч

Применение правил фильтрации также положительно влияет на результат, но есть ряд вопросов, требующих дальнейшего изучения. Вместе с увеличением точности снижается отзыв. Для решения этой проблемы целесообразно рассмотреть другие критерии выбора правила, чтобы найти подходящие экспериментальные значения граничных параметров для специфичности правила и поддержки аспектов-кандидатов для достижения минимального снижения отзыва.

4. Выводы и будущая работа

В статье рассматриваются подходы к явному извлечению аспектов и классификации настроений. Алгоритм, основанный на методе опорных векторов, показал лучшее качество. Алгоритм на основе k ближайших соседей показал худший результат. Результаты находятся на уровне средних результатов, представленных в треке анализа настроений. Алгоритм на основе SVM с использованием в качестве признаков нормализованной леммы и синтаксических связей показал наилучшие результаты на треке. В попытках выделить аспекты можно сказать, что самый простой подход показывает сравнимые с остальными результаты. Использование правил фильтрации

для повышения точности при уменьшении полноты. В связи с этим необходимо отдельно оценивать влияние граничных параметров на результат.

В качестве будущей работы можно выделить такие направления, как: применение полууправляемых методов генерации правил для снижения трудозатрат, использование активных методов обучения, построение системы визуализации генерации правил, способной обеспечить процесс взаимодействия с экспертами. . Также расширение модели правила фрагмента может дать новые выразительные возможности.

Рекомендации

1. Go A., Huang L., Bhayani R. (2009), Классификация настроений в Твиттере с использованием дистанционного наблюдения, Отчет о проекте CS224N, Стэнфорд.
2. Кеннеди А., Инкпен Д. (2006), Классификация настроений рецензий на фильмы с использованием контекстных сдвигателей валентности, Computational Intelligence, vol. 22(2), стр. 110–125.
3. Li S., Lee SYM, Chen Y., Huang C.-R., Zhou G. (2010), Классификация настроений и смена полярности, Материалы 23-й Международной конференции по компьютерной лингвистике (Coling, 2010), Пекин, Китай, стр. 635–643.
4. Лю Б. (2012), Анализ настроений и изучение мнений, Морган и Клейпул. Издатели.
5. Лукашевич Н., Блинов П., Котельников П., Рубцова Ю., Иванов В., Тутубалина Е. (2015), SentiRuEval: Тестирование систем объектно-ориентированного анализа настроений на русском языке.
6. Накагава Т., Инуи К., Курохаши С. (2010), Классификация настроений на основе дерева зависимостей с использованием CRF со скрытыми переменными, Ежегодная конференция Североамериканского отделения ACL, 2010 г., Лос-Анджелес, США, стр. 786. –794.
7. Панг Б., Ли Л., Вайтьянатан С. (2002), Недурно? Классификация настроений с использованием методов машинного обучения, Труды EMNLP, Филадельфия, Пенсильвания, США, стр. 79–86.
8. Парсер Томита: https://tech.yandex.ru/tomita/ 9. Васильев В.Г. (2011), Классификация и выделение фрагментов в текстах на основе логических правил. Электронные библиотеки: передовые методы и технологии, Электронные коллекции RCDL'2011, Воронеж, стр. 133–139.
10. Васильев В.Г., Давыдов С.У. (2013), Классификация настроений комбинированным подходом. Компьютерная лингвистика и интеллектуальные технологии: Материалы международной конференции «Диалог 2013» . Режим доступа: www.dialog-21.ru/digests/dialog2013/materials/pdf/VasilyevVG.pdf 11. Заки М.Дж., Партасарати С., Огихара М. , Ли В. (1997), Новые алгоритмы для быстрого обнаружения правил ассоциации, Труды Третьей международной конференции по обнаружению знаний в базах данных и интеллектуальном анализе данных, Ньюпорт-Бич, Калифорния, США, стр. 283–286.