

Правительство Российской Федерации

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования**

**Национальный исследовательский университет
"Высшая школа экономики"**

Факультет филологии

Направление подготовки 035800.68

«Фундаментальная и прикладная лингвистика»

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

На тему «Выявление скрытой оценки»

Студент группы № 72/л

Сутормина Алиса Юрьевна

Руководитель магистерской
диссертации

к.ф.н., доцент

Толдова Светлана Юрьевна

Москва, 2014 г.

Оглавление

Введение	3
Глава 1. Задачи и методы в области автоматизированного извлечения мнений.....	6
1.1. Введение	6
1.2. Задачи автоматизированного извлечения мнений.....	7
1.3. Методы, применяемые для автоматизированного извлечения мнений	10
1.4. Методы составления словарей оценочной лексики	15
1.5. Выводы.....	18
Глава 2. Создание макета системы извлечения явной и скрытой оценки.	20
2.1. Введение	20
2.2. Построение и тестирование системы извлечения явной оценки	21
2.2.1 Составление оценочного словаря	21
2.2.2 Составление списков слов, участвующих в выражении оценки	23
2.2.3 Построение правил извлечения оценки.....	25
2.2.4 Тестирование системы	32
2.3. Построение модуля извлечения скрытой (опосредованной) оценки.....	33
2.3.1 Выявление правил извлечения скрытой оценки.....	33
2.3.2 Тестирование системы, дополненной модулем извлечения скрытой оценки.	43
2.4. Выводы.....	43
Заключение.....	44
Приложение 1. Список конструкций, переносящих оценку	45

Введение

Извлечение мнений (*opinion-mining*) – активно исследуемая на сегодняшний день и довольно обширная задача, лежащая в сфере компьютерной лингвистики, в частности – такой ее области, как извлечение данных из текста. Востребованность автоматизированных программных инструментов, позволяющих выявлять оценочные характеристики восприятия объекта или явления целевой аудиторией, в политической, экономической, социологической и рекламной сферах обуславливает актуальность тематики данной работы.

Существующие и разрабатываемые системы извлечения мнений различаются не только используемыми в их архитектуре технологиями и методами, но и в соответствии с широким кругом решаемых с их помощью подзадач – таких, как определение общей тональности текста, выделение оценки относительно известного или неизвестного объекта, разграничение субъективных и объективных высказываний и т.д.

В настоящей работе мы будем рассматривать конкретную подзадачу области извлечения мнений, а именно – выделение оценки и связывание ее с оцениваемым объектом. Один из распространенных методов решения этой подзадачи – составление лексико-грамматических шаблонов и словарей оценочной лексики – будет использован в данной работе.

Следует заметить, что вариативность выражения оценки в русском языке вызывает определенные трудности в области ее автоматизированного анализа; такие факторы, как многозначность оценочных слов и многообразие синтаксических конструкций в значительной степени влияют на точность и полноту результатов работы извлекающих оценку систем.

Отдельно стоит выделить случаи, когда объект оценивается не явно (эксплицитно), а опосредованно (имплицитно); в ряде работе [Ермаков, Киселев; 2005] такие ситуации также называются эмоционально-коннотативными. Такие ситуации нередко встречаются в авторских, индивидуализированных текстах, в особенности в рамках политического дискурса, что в значительной мере обусловлено его прагматикой [Демьянков; 2002] – в частности, в тех случаях, когда автор высказывания по той или иной причине избегает явной оценки обсуждаемого объекта (как показывают примеры, этой причиной нередко является желание автора сделать текст более «объективным»). К примеру, в предложениях «президент борется с коррупцией», «Госдума одобрила сомнительный законопроект» или «Девушки организовали отвратительную акцию» выделенный курсивом объект

оценивается при помощи его связи с другим явно оцененным или имеющим определенную коннотацию объектом. Анализ такого рода конструкций методом шаблонов считается рядом исследователей довольно сложным, требующим глубокого семантического анализа [Кравченко; 2012], и не используется в большинстве известных систем. Это в значительной степени влияет на полноту результатов их работы, так как выделенная оценка часто не связывается с непосредственно оцениваемым автором текста объектом.

Именно такого рода ошибки и определяют проблематику данной исследования. Проанализировав переносящие оценку конструкции и обобщив их в виде шаблонов, мы предпримем попытку улучшить качество работы разрабатываемого нами макета системы извлечения мнений при помощи модуля извлечения скрытой оценки. Кроме того, чтобы увеличить степень объективности тестирования, так или иначе зависящего от интерпретации экспертом оценки в примерах приведенного типа, мы осуществим разметку тестового корпуса по объектам, положительное или отрицательное отношение автора к которым ясно из текста, параллельно тремя не относящимися к исследованию экспертами, оставив в качестве результирующей ту разметку, с которой согласились минимум два эксперта из трех. Полученные таким способом в результате тестирования данные наиболее полно говорят о целесообразности введения модуля извлечения скрытой оценки, сводя к минимуму субъективность ее восприятия.

Таким образом, объектом нашего исследования является скрытая, или опосредованная оценка объекта, а предметом – способы ее выражения в тексте, которые мы опишем при помощи шаблонов.

Исходя из вышесказанного, цель настоящего исследования – выявить лексико-грамматические шаблоны извлечения имплицитной оценки и доказать, что их использование в системе анализа тональности текста способно качественно улучшить результаты ее работы.

Указанная цель подразумевает решение следующих задач:

1. Обзор использующихся методов извлечения оценки и ее связывания с объектом;
2. Создание системы извлечения эксплицитной (явной) оценки на основе лексико-грамматических шаблонов, использующихся в современных системах, и списка оценочных слов, подставляемых в данные шаблоны;

3. Измерение точности и полноты результатов применения используемых шаблонов;
4. Анализ способов выражения имплицитной (опосредованной) оценки, составление лексико-грамматических шаблонов ее извлечения и установления связи с релевантным объектом;
5. Измерение точности и полноты результатов работы системы, имеющей модуль извлечения имплицитной оценки, и сравнение этих результатов с результатами работы системы, использующей только правила извлечения явной оценки.

К способам достижения поставленной цели можно отнести, в первую очередь, проведение экспериментов. В качестве методик будет использован корпусный анализ – как проверка определенных предположений при помощи Национального корпуса русского языка, так и анализ данных размеченного нами корпуса политических блогов объемом 200000 слов. Для решения поставленных задач в первой главе исследования будет проведен обзор различных направлений и методов извлечения оценки из текста. Во второй главе, во-первых, будут рассмотрены и апробированы использующиеся в большинстве систем шаблоны и проведен эксперимент с оценкой качества результатов их использования; во-вторых, будут выявлены и классифицированы шаблоны опосредованного (имплицитного) выражения оценки, измерена точность и полнота результатов и сделаны выводы о целесообразности их введения в архитектуру системы извлечения мнений.

Глава 1. Задачи и методы в области автоматизированного извлечения мнений

1.1. Введение

Извлечение мнений – это область компьютерной лингвистики, которая занимается выявлением оценок и мнений, представленных в тексте. Устоявшиеся англоязычные термины для определения данной сферы – *opinion mining* и *sentiment analysis*.

Термин «*opinion mining*», впервые введенный в работе [Dave, Lawrence, Pennock; 2003], изначально понимался более узко – как выявление и анализ суждений, касающихся определенных сторон какого-либо предмета. Словосочетание «*sentiment analysis*» появилось сразу в нескольких работах – сначала в [Das, Chen; 2001] и [Tong; 2001], а затем и в множестве других. Большинство авторов, упоминая *sentiment analysis*, фокусировались на классификации отзывов на положительные и отрицательные, из-за чего данный термин на какое-то время стали относить только к данной подзадаче; тем не менее, на данный момент он понимается более широко – как совокупность методов автоматического выделения из текста мнения, тональности и субъективной оценки [Pang, Lee; 2008].

На данный момент широкая интерпретация обоих терминов охватывает примерно одну и ту же область, которая в свою очередь является более узкой сферой анализа субъективности [Pang, Lee; 2008], и позволяет более-менее равноценно использовать оба словосочетания для обозначения данного раздела компьютерной лингвистики.

Русскоязычный термин «определение тональности текста», значение которого практически идентично первоначальному значению *sentiment analysis*, довольно распространен в отечественной литературе; несмотря на то, что *тональность* является лишь одной из характеристик субъективности, этот аспект является наиболее исследуемым (возможно, в силу относительной простоты реализации методов его изучения).

В последнее время принято говорить об извлечении мнений (наиболее близкий эквивалент словосочетания «*opinion-mining*»), или извлечении оценки. Сфера употребления этих терминов несколько отличается от «определения тональности», которая в большей степени подразумевает классификацию текста или его фрагмента в соответствии с его общей эмоциональной направленностью; «извлечение мнений» и «извлечение оценки» актуально в случае, когда требуется найти в тексте субъективное высказывание, определить полярность оценки и связать ее с релевантным объектом. В данной работе мы будем использовать именно эти термины, как в наибольшей степени отражающие проблематику исследования.

Исходя из вышесказанного, применимость того или иного термина в любом случае связана с конкретной задачей; подробный обзор задач исследуемой области будет произведен в следующем разделе.

1.2. Задачи автоматизированного извлечения мнений

Как задачи, так и методы в области извлечения мнений в значительной степени зависят от типа анализируемого текста. Самым распространенным материалом для анализа являются тексты, находящиеся в свободном доступе в сети Интернет; эти тексты очень многообразны, неоднородны и различаются по множеству важных для их обработки параметров.

На наш взгляд, об извлечении оценки как таковой целесообразно говорить только в отношении авторских, индивидуализированных текстов – таких, как блоги, записи в твиттере, сообщения на форумах, отзывы и т.д. Индивидуализированные тексты чаще всего анализируются с точки зрения выявления отношения автора к какому-либо объекту. Что касается обработки информативных текстов, субъективность которых минимальна (к примеру, новостных), то в рамках исследуемой области мы можем выделить следующие задачи их анализа:

1. Определение общей тональности текста (что в определенном смысле является предсказанием возникающей при чтении данного текста эмоциональной реакции);
2. Определение тональности контекста, в котором употребляется интересующий нас объект (т.е. выявление «имиджа» какой-либо персоны, формируемого СМИ; сюда также могут относиться такие «политические» задачи, как определение степени поддержки кандидата или предсказание результатов выборов, представленные в работах [Thomas et al, 2006] и [Kim and Novy, 2007]);
3. Выделение прямой речи и цитат, в которых может производиться оценка объекта или ситуации.

Индивидуализированные тексты, в свою очередь, можно разделить на две различные для их анализа подгруппы:

1. Ограниченный класс текстов, являющихся отзывами о каком-либо продукте (к примеру, фильме, книге или товаре). Такие тексты обладают определенными чертами, облегчающими их анализ – заданностью объекта оценки, наличием некоторых типичных для отзывов конструкций, относительной предсказуемостью;

2. Остальные авторские тексты сети Интернет – блоги, твиты, сообщения на форумах и в социальных сетях; несмотря на то, что тематика этих текстов может быть известна, выраженная в них оценка очень многообразна и часто непредсказуема.

Кроме того, тексты последней упомянутой группы могут различаться по длине и нормативности – к примеру, записи в блогах в большей степени структурированы, в то время как сообщения в твиттере отличаются небольшой длиной, высокой компрессией, активным использованием эмодзи, отступлением от литературных норм и т.д. В связи с тем, что морфологический и синтаксический анализ таких сообщений затруднителен, к ним чаще применяются статистические методы, тогда как диапазон инструментов анализа более структурированных индивидуализированных текстов может включать их полноценный лингвистический анализ.

Учитывая вышесказанное, для апробации методов выявления оценки и ее связи с объектом мы считаем наиболее целесообразным использовать в качестве обучающего и тестового материала тексты блогосферы, отличающиеся не только ярко выраженной субъективностью, но и значительной вариативностью формулировок оценочных высказываний.

В рамках анализа индивидуализированных текстов обычно рассматривается классическая оценочная ситуация, в которой автором высказывания производится суждение, выражающее его мнение о каком-либо объекте. В качестве элементов такой ситуации выделяют:

- субъект оценочного высказывания, то есть тот, кто высказывает некоторое субъективное мнение (в большинстве случаев – автор текста);
- оценочное высказывание, где выражена оценка некоторого объекта;
- объект оценки, то есть тот объект или явление, о котором высказывается оценка;
- полярность оценки (положительная или отрицательная; некоторые также выделяют нейтральную).

Выделяют также «необязательные» элементы оценочной ситуации:

- Параметры объекта, подвергающиеся оценке (чаще всего, параметры объекта известны в случае, если речь идет об анализе отзывов и объект оценки

известен; оценка разных параметров одной и той же продукции может иметь разную «полярность»);

- В ситуации оценивания также неявно присутствует тот, кто воспринимает эту оценку, или «потребитель» оценки.

Исходя из вышесказанного, попытаемся еще раз представить все многообразие направлений области извлечения мнений, перечислив подзадачи, которые требуется или не требуется решать в зависимости от поставленной цели:

- Выделение оценочного (субъективного) высказывания.

Выделять оценочное высказывание не требуется, если нас интересует общая тональность текста (как случае с информативными текстами, так и в случае с отзывами, которые заведомо полностью являются оценочными (работа [Pang and Lee, 2005], [Клековкина, Котельников; 2012]); с другой стороны, если нас интересует оценка конкретного объекта, выделять такие высказывания можно как в информативных текстах (цитаты в новостях), так и в индивидуализированных текстах (оценочные высказывания в блогах). Разграничению субъективной и объективной информации посвящены такие работы, как [Wiebe, 2000]; [Wiebe et al., 2004]; [Wiebe and Riloff, 2005].

- Определение полярности оценки.

Такая задача, как определение полярности оценки, может быть опущена в том случае, если принципиально важно только отделить субъективную информацию от объективной; в других случаях речь может идти как о разделении оценочных высказываний на 2 класса (позитив и негатив), так и на три класса (позитив, негатив, нейтральный), а также об определении интенсивности оценки по более широкой шкале.

- Выделение объекта оценки.

Выделять объект оценки не требуется в том случае, если определяется общая тональность текста, или если объект оценки заранее известен (к примеру, при анализе отзывов); в остальных случаях может выделяться либо любой оцениваемый автором объект, либо какая-либо их группа (например, персоны), либо только один объект, оценка которого нас интересует (например, при анализе имиджа определенной персоны в СМИ). Выделение объекта оценки как отдельная задача представлено в работах [Hurst and Nigam; 2004]; [Gamon and Aue; 2005]; [Hu and Liu; 2004]; [Popescu and Etzioni; 2005]; [Kim and Novy; 2006]; [Kobayashi et al.; 2007].

- Выделение субъекта оценки.

Субъект оценки выделяется, когда речь идет либо о наличии субъективных высказываний в информативный текст (цитаты в новостях), либо о цитировании в индивидуализированных текстах; в других случаях обычно субъектом оценки по умолчанию считается автор текста ([Пазельская, Соловьев; 2011]). Следует также упомянуть исследования, направленные исключительно на определение субъекта и объекта оценки, представленные в работах [Kim, Novy; 2006], [Kobayashi; 2007].

- Определение длины фрагмента, тональность которого анализируется.

В соответствии с различными целями может анализироваться тональность как всего текста [Mishne and Glance; 2006], [Mishne; 2005], [Leshed and Kaye; 2006], так и одного предложения; кроме того, могут «суммироваться» все встретившиеся в одном или нескольких текстах оценки относительно интересующего нас объекта.

В рамках данного исследования нами будет разработан макет «классической» системы извлечения мнений, задачей которой будет:

1. Найти все оценочные высказывание в тексте (за основную единицу анализа будет взято предложение);
2. Определить полярность оценки и связать ее с оцененным объектом.

Что касается объекта оценки, то будут учитываться абсолютно все объекты, как-либо оцененные автором, но суммироваться эта оценка не будет. В качестве полярности оценки мы будем различать положительную и отрицательную, то есть разбивать оцененные объекты только на два класса; нейтральная оценка будет приравниваться нами к отсутствию оценки.

Таким образом, задачи в области извлечения мнений настолько многочисленны и разнообразны, что чаще всего требуют использования совершенно разных методов их решения. Наиболее часто применяемые методики и алгоритмы будут рассмотрены ниже.

1.3. Методы, применяемые для автоматизированного извлечения мнений

Дифференцирование методов автоматизированного извлечения оценки коррелирует с глобальным противопоставлением двух используемых в компьютерной лингвистике подходов – подхода, основанного на правилах (*rule-based approach*) и подхода, основанного на статистическом анализе данных (*data-based approach*), в большинстве

случаев так или иначе связанного с машинным обучением. Использование правил также требует наличия словаря оценочной лексики (который, в то же время, может использоваться и при машинном обучении). Данные подходы также могут использоваться в различных комбинациях.

Метод классификации текстов по тональности с помощью машинного обучения является очень распространенным при решении данной задачи. Примером такого подхода могут служить системы, принявшие участие в соревновании по sentiment analysis в рамках семинара РОМИП в 2011 году. Описание данных систем представлено в работах [Клековкина, Котельников; 2012] и [Pak, Paroubek; 2012].

Такой метод предполагает использование различных статистических алгоритмов классификации, наиболее распространенными из которых являются метод опорных векторов, алгоритм Байеса, модель максимальной энтропии; в последнее время стал также популярен метод условных случайных полей [Антонова, Соловьев; 2013].

При данном подходе, во-первых, необходимо иметь обучающий корпус, а, во-вторых, определить пространство признаков, используемых для классификации. В качестве классифицирующих параметров могут использоваться:

- Все лексемы текста;
- Отдельные списки лексем по каждой частеречной категории ([Wiebe; 1999], [Bruce, Wiebe; 2000]);
- Только оценочные предикаты (чаще всего – прилагательные ([Pang, Lee, Vaithyanathan; 2002]), определяющие ориентацию оценки, а также списки предикатов мнения) ;
- Списки лексем, диагностирующих, что данный контекст субъективный, а также лексем, определяющих ориентацию оценки;
- N-граммы: могут использоваться как словарные биграммы (которые также эффективны в борьбе с отрицанием) ([Dave; 2006], [Aue, Gamon; 2005], [Wiebe, 2001], [Riloff ; 2006]), так и символьные. Их использование целесообразно для определения слабоструктурированных текстов, полноценный анализ и лемматизация которых не очень эффективна – к примеру, сообщений в твиттере; кроме того, при использовании символьных N-грамм учитываются эмодзи;
- Списки лексико-грамматических шаблонов (к примеру, предикат с моделью управления);
- Списки кортежей (например, топик – оценка – ориентация оценки);

- Можно также учитывать отрицание, то есть использовать в качестве параметров как просто лексемы, так и лексемы с отрицанием

При классификации может учитываться как наличие или отсутствие параметра, так и его вес или полярность. Может быть учтена частота параметра в положительных контекстах в обучаемом корпусе, в отрицательных контекстах и в нейтральных. Кроме того, часто используется взаимная информация, учитывающая совместную встречаемость в некотором окне с лексемами, дающими высокую точность.

Перечисленные параметры могут быть дополнены контекстной информацией ([Wieber et al., 2001]; [Andreevskia et al., 2007]), а также признаками, отражающими особенности синтаксической структуры ([Gamon, 2004]). Существуют работы, в которых учитывалась и позиция оценочного предложения в тексте ([Taboada and Grieve, 2004]).

Несмотря на относительную универсальность таких методов, среди многих авторов популярно утверждение о целесообразности использования «предметнозависимого» классификатора, то есть классификатора, который будет обучаться и тестироваться на текстах одной предметной области. Такой подход, действительно, более эффективен, так как учитывает многозначность оценочной лексики в зависимости от тематики и типа текста.

Противоположный подход, основанный на правилах и словарях, характеризуется хорошей точностью при решении задачи выявления оценки и ее связи с объектом. На наш взгляд, именно этот метод целесообразно использовать в рамках настоящего исследования, как в наибольшей степени соответствующей решаемой задаче.

Данная методика основана на составлении правил с использованием словарей тональной лексики и подразумевает составление шаблонов для поиска тонально окрашенных фрагментов текста. Такой подход подробно описан в работе [Пазельская, Соловьев; 2011].

Для того, чтобы увеличить полноту работы системы, основанной на правилах, целесообразно осуществлять определенную предобработку текста – разрешать анафорические ссылки, разбивать сложные предложения на простые (в зависимости от типа придаточных некоторые из них могут включаться в родительское), приводить причастные и деепричастные обороты к определенным шаблонам, и только потом извлекать из получившейся структуры участников ситуации [Пазельская, Соловьев; 2011].

При непосредственном анализе первым шагом является морфологический анализ текста, а также присвоение тональности отдельным лексемам в соответствии с данными

словарей тональной лексики. Далее следует либо полноценный, либо поверхностный синтаксический анализ и набор правил объединения оценок в цепочки (сочетания оценочных слов с инверторами, шифтерами, оценочными словами других классов). В последнюю очередь срабатывает правило выделения объекта оценки. [Пазельская, Соловьев; 2011].

Для составления шаблонов выделения оценки необходимо подробно классифицировать оценочные ситуации и входящие в них пропозиции; довольно полная их классификация представлена в работе [Ермаков, Киселев, 2005], где авторы выделяют следующие классы ситуаций:

1. Явная тональная характеристика

Ситуации этого класса выражаются несколькими пропозициями:

- Логическая пропозиция полная: Петров – *плохой* руководитель; власть становится *защитницей преступности*; политика власти - *дрянь*. Тональность выражается именной группой, образуемой существительным.
- Логическая пропозиция свернутая с существительным: *гениальный авантюрист* Петров; *робкая политика* президента, *циничность* власти. Тональность выражается именной группой, образуемой существительным.
- Логическая пропозиция свернутая с прилагательным: *нерешительный* президент; президент *нерешителен*. Тональность выражается прилагательным.
- Свернутая логическая пропозиция в составе событийной, отражающая оценку события, в котором целевой объект выступает в роли протагониста: Петров *бездумно* согласился; президент принял *авантюрное* решение; власть *обнаглела*. Тональность может выражаться наречием при глаголе, прилагательным при событийном существительном, самим глаголом.

2. Прямая эмоционально-коннотативная характеристика

Класс эмоционально-коннотативных ситуаций, отражающих отношение целевого объекта к эмоционально-коннотативным сущностям, их оценку целевым объектом, или наоборот - отношение этих сущностей к объекту, оценку объекта ими. Выражается событийными пропозициями: президент *борется с преступностью*, *народ* выносит *осуждение* власти. Выделяются участники в следующих ролях:

- *Субъект* - активный участник, в приведенных примерах - "президент" и "народ";
- *Объект* - пассивный участник ситуации, в приведенных примерах - "преступность" и "политика власти";
- *Предикат* - глагол или существительное, выражающее отношение *Субъекта* к *Объекту* (в приведенных примерах - "бороться с" и "осуждение").

Типовые пропозиции, которыми выражаются ситуации этого класса, следующие:

- Событийная пропозиция полная или свернутая, в которой роль *Субъекта* занимает целевой объект: власть *борется с олигархами*; президент ведет *борьбу с коррупцией*, *борьба* президента за *права народа*. Тональность складывается из семантики именной группы в роли *Объекта* и семантики *Предиката* по принципу "положительное отношение к положительному - позитив" и наоборот. Если эмоциональный коннотат *Объекта* или *Предиката* не определен, тональность считается нейтральной (президент встретился с *олигархами*, Иванов *борется с сорняками* на даче).
- Событийная пропозиция полная или свернутая, в которой роль *Объекта* занимает целевой объект: *олигархи испугались* президента, *страна* выражает *недоверие к власти*, *ненависть преступников к власти*. В случае, если семантика *Субъекта* имеет положительный эмоциональный коннотат, общая тональность складывается по тому же принципу, что и в (1). Если же семантика *Субъекта* имеет отрицательный коннотат, то общая тональность не определена: *олигархи полюбили* президента (да мало ли за что?).

3. Ассоциированный эмоциональный коннотат

Класс эмоционально-коннотативных ситуаций, фигурирующих в одном предложении с целевым объектом, но не связанных с ним напрямую (в ряде случаев эту связь просто не удастся идентифицировать средствами автоматического анализа текста). Выражаются событийными пропозициями: *обнищание пенсионеров*, *повысить зарплату*, *рождаемость падает*, *пристраститься к пиву*. Выделяются участники в следующих ролях:

- *Участник* - участник, на состояние которого влияет событие;
- *Предикат* - событие, которое влияет на *Участника*.

Тональность складывается из семантики именной группы в роли *Участник* и семантики *Предиката* по принципу "хорошо для хорошего - позитив" и наоборот. Если эмоциональный коннотат *Участника* или *Предиката* не определен, тональность считается нейтральной.

Ситуации первого типа часто подвергаются анализу и фигурируют во многих исследованиях, в то время как ситуации второго и третьего типа практически не охватываются существующими шаблонами, так как, по мнению ряда исследователей, их анализ довольно труден [Кравченко; 2012].

В данной работе мы попытаемся не только создать шаблоны для анализа такого рода ситуаций, но и выявить и классифицировать большую часть выражающих их пропозиций. Говоря о классификации Ермакова и Киселева, мы будем придерживаться введенного нами разделения оценки на эксплицитную (явную) и имплицитную (скрытую) – в данном случае к эксплицитной оценке будет относиться явная тональная характеристика, а к имплицитной – как прямая, так и ассоциированная эмоционально-коннотативная ситуация. Кроме того, мы попытаемся выявить и классифицировать дополнительные шаблоны, описывающие способы выражения имплицитной оценки.

Итак, основные этапы использования метода, основанного на правилах, заключаются, во-первых, в формировании лексико-грамматических шаблонов и правил их срабатывания и, во-вторых, в составлении словарей тональной лексики (для подставления в шаблоны). Методы составления словарей тональной лексики (а также классы входящих в них слов) будут рассмотрены в следующем разделе.

1.4. Методы составления словарей оценочной лексики

Лексику, так или иначе определяющую тональность текста, в самом обобщенном виде можно разделить на следующие классы:

- Списки слов, явно выражающие позитивную или негативную оценку (в первую очередь, оценочные прилагательные и существительные типа *молодец*);
- Интенсификаторы (например, *очень*);
- Инверторы (слова, инвертирующие оценку – *не, без* и т.д.);
- Предикаты мнения и оценки;
- Предикаты, содержащие имплицитную или эксплицитную оценку в зависимости от роли объекта оценки в ситуации;

- Амбивалентная лексика (оценочные лексемы, ориентация которых зависит от контекста);
- Лексика, обладающая положительной или отрицательной коннотацией (к примеру, *коррупция*).

В зависимости от конкретной цели анализа и желаемой полноты могут быть сформированы и другие классы лексики. К примеру, в работе [Пазельская, Соловьев; 2011] были выделены лексемы следующих классов:

1. Прилагательные (позитивные, негативные и усилительные);
2. Наречия (позитивные, негативные и усилительные);
3. Существительные (позитивные, негативные, потенциально позитивные и потенциально негативные, отглагольные негативные и отглагольные позитивные);
4. Глаголы:

- 1 и 2 класс — негативные и позитивные глаголы, определяющие тональность объекта в зависимости от окружения и независимо от роли (негативные *уносить, освободить от*; позитивные *защищать, болеть за*);

- 3 и 4 класс — негативные и позитивные глаголы, определяющие тональность объекта независимо от окружения, но в зависимости от его роли (*сдаться* и *проиграть* – субъект негативный, объект позитивный; *обуздать* и *повергнуть* – субъект позитивный, объект негативный);

- 5 и 6 класс — негативные и позитивные глаголы, определяющие тональность объекта в зависимости от окружения и роли (негативные – *жаловаться, испугаться*; позитивные – *окупаться, согреться*);

- 7 и 8 класс — чисто негативные и чисто позитивные глаголы, определяющие тональность объекта вне зависимости от его роли и окружения (позитивные *расследовать* и *улучшить*, негативные *грабить* и *злоупотреблять*);

- 9 класс — глаголы, отождествляющие тональность объекта и субъекта (*являться, олицетворять, относиться*);

1. Глагольные коллокации (позитивные и негативные);
2. Неглагольные коллокации (позитивные, негативные и усилительные);

Состав оценочного словаря и классификация входящей в него лексики в большой степени определяется конкретной задачей. В указанной работе авторы ставили перед

собой задачу анализа тональной окраски новостей, в то время как нами будут анализироваться индивидуализированные тексты, где интерес представляет именно авторская точка зрения на релевантный объект, а не восприятие читателем тональности фрагмента текста. Данная цель значительно сокращает желаемый объем словаря оценочной лексики, исключая из его состава такие слова, как «согреться» или «вмешаться», оценочная функция которых довольно сомнительна. Кроме того, в настоящей работе будет проведена другая классификация оценочной лексики, в наибольшей степени отвечающая требованиям разрабатываемых в рамках исследования шаблонов.

Среди непосредственных методов составления оценочных словарей можно выделить следующие направления:

1. Составление списков тональной лексики вручную.

В вышеупомянутой работе [Пазельская, Соловьев; 2011] были составлены экспертные словари размером 15000 лексем. Данный подход особенно эффективен в рамках решения конкретной задачи извлечения тональности из текстов конкретных тематик, но является очень трудоемким.

2. Использование готовых словарей и списков.

Данный метод более актуален для английского языка, т.к. для русского языка на данный момент существует немного готовых словарей или списков. Список тональной лексики, неразмеченной на позитив и негатив, доступен на сайте РОМИП; кроме того, в Семантическом словаре русского языка под редакцией Н.Ю. Шведовой один из томов содержит существительные, размеченные в соответствии с положительной или отрицательной коннотацией. Также может использоваться семантическая разметка Национального корпуса русского языка.

3. Автоматические и полуавтоматические методы

–Извлечение лексики из корпусов текста

- Классификация лексики при наличии небольшой размеченной выборки или базового списка с помощью стандартных алгоритмов машинного обучения;
- Выделение синтаксических шаблонов типа «прилагательное – существительное» и «наречие – глагол»;

- Использование информации о частотности слов в «субъективных» и «объективных» текстах;

- Извлечение прилагательных из сочинительных конструкций (данный метод, основанный на предположении, что в сочинительной конструкции участвуют семантически близкие прилагательные, описан в работе [Hatzivassiloglou, McKeown; 1997];

- Бутстрэппинг, или «метод снежного кома». Возможны различные вариации данного метода, но суть его заключается в том, чтобы, имея небольшое множество слов и конструкций, диагностирующих тональность с хорошей точностью, получить обучающую выборку, а затем увеличивать полноту за счет статистических методов. К примеру, на основе базового списка (минимум слов, дающих максимальную точность) составляются ручные правила, работа правил проверяется на небольшом размеченном массиве и используются стандартные методы классификации для выделения новых слов и конструкций (т.е. выделение новых признаков); в качестве параметров берутся те слова, которые дают высокую точность при классификации (т.е. правильно классифицирующие тексты с точностью, большей некоторого порога). Применение данного метода описано в работах [Riloff, Wieber; 2003], [Esuli, Sebastiani; 2005].

–Использование WordNet (для английского языка)

- Оценка ориентации лексемы по тому, какова длина ее пути до слова – прототипа (например, *good* или *bad*) [Kamps, Marx, Mokken; 2004];

- Классификация синсетов [Andreevskaia, Berger; 2006], [Esuli, Sebastiani; 2007].

–Использование FrameNet (например, выделение глаголов с экспериенцером)

В рамках данного исследования нами будут экспертно размечены и объединены существующие источники субъективной лексики; кроме того, часть слов будет получено из размеченного нами корпуса текстов, по тематике совпадающим с тестовым корпусом.

1.5. Выводы

Таким образом, мы рассмотрели широкий круг задач, связанных с проблемой определения тональности текста, а также совокупность методов, использующихся для решения данной задачи. Резюмируя все вышесказанное, можно сделать несколько заключений:

- Определение тональности текста – это очень широкая область, которая включает большое количество подзадач, требующих совершенно различных алгоритмов и методов решения;

- Метод определения тональности следует выбирать с учетом конкретной задачи и типа анализируемого текста; тональная лексика многозначна и зависит от контекста, поэтому «универсальный» классификатор будет показывать гораздо меньшую точность, чем предметно ориентированный;

- Большинство работ показывают, что методы машинного обучения (классификации) скорее применяются для обработки либо коротких неструктурированных текстов (твитов), либо заведомо классифицируемых (отзывы), в то время как для анализа длинных и структурированных текста разработчики коммерческих систем предпочитают применять правила; в то же время, в иностранной литературе все большую популярность приобретают гибридные подходы;

- Для русского языка, в отличие от английского, существует очень мало общедоступных списков и словарей тональной лексики, и составление таких списков всегда является первоочередной проблемой при решении данной задачи, поэтому требуется вести работу и в этом направлении.

В рамках настоящего исследования нами будет разработана система, которая будет обучаться и тестироваться на индивидуализированных структурированных текстах однородной тематики; основу системы будет составлять продукционный классификатор, состоящий из лексико-грамматических шаблонов, правил и списка оценочных слов для подстановки в шаблоны. Система будет включать как «стандартные» правила извлечения эксплицитной оценки, так и модуль, извлекающий имплицитную оценку, результат применения которого будет оценен впоследствии.

Глава 2. Создание макета системы извлечения явной и скрытой оценки.

2.1. Введение

Итак, в рамках практической части данной работы будет создан макет системы извлечения оценки, основанной на шаблонах и правилах. Помимо правил извлечения эксплицитной оценки, в системе также будет присутствовать модуль извлечения имплицитной оценки; таким образом, нами будет подсчитана точность и полнота результатов работы системы как с данным модулем, так и без него, в результате чего будут сделаны выводы о целесообразности его использования.

Сформулируем некоторые требования, предъявляемые к создаваемой системе:

1. На вход принимается текст, разбитый на предложения;
2. Результатом работы системы должен быть список объектов, оцененных автором текста (в каждом конкретном предложении) и полярность оценки, соответствующей каждому из объектов; полярность оценки может быть только положительной или отрицательной.
3. Разрешение анафоры и кореференции представляет собой отдельную задачу и выходит за рамки данной работы; таким образом, мы рассматриваем упоминание и оценку объекта в каждом конкретном предложении.
4. Объектом оценки могут быть персоны, местоимения, именные группы.

В качестве инструмента для создания системы извлечения оценки мы будем использовать продукт Tomita-парсер компании Яндекс. Данный анализатор включает в себя три стандартных лингвистических процессора – токенизатор, сегментатор и морфологический анализатор. Ограниченный список грамем (грамматических характеристик), полученных в процессе морфологического анализа, в дальнейшем используется при написании грамматики. Важным элементом работы анализатора также является словарь ключевых слов, которые впоследствии используются в правиле.

Далее парсер позволяет по написанным экспертом шаблонам (правилам контекстно-свободной грамматики) выделять из текста разбитые на поля цепочки слов. Таким образом, в рамках данного исследования наша задача состоит в составлении правил, в котором одни цепочки будут проинтерпретированы как объекты оценки, а другие – как выражение оценки положительной или отрицательной полярности.

Для формирования тестового и обучающего (т.е. использующегося для выявления шаблонов, составления списков оценочных слов и т.д.) корпусов нами были использованы

тексты политических блогов, взятые с сайтов livejournal.com и echo.msk.ru за период с октября 2012 г. по январь 2013 г. Блоги были субъективно отобраны по популярности и соответствии политической тематике. Выбор тематики обусловлен многими факторами – большей длиной и литературностью текстов, обилием оценочных высказываний (при этом не всегда явных), вариативностью выражения оценки. Размер обучающего корпуса составил 200000 слов (350 блоговых записей); размер тестового корпуса – 100 записей (около 50000 слов). В обучающем корпусе нами были выделены объекты оценки, выражающие оценку слова (по частям речи), а также слова, на наш взгляд переносящие оценку, т.е. служащие выражению имплицитной оценки. Для получения более достоверных данных тестовый корпус размечался параллельно тремя экспертами; при этом в нем были выделены только объекты, отношение к которым ясно из текста и может быть охарактеризовано как положительное и отрицательное. В качестве конечного варианта корпуса был выбран список объектов, которые выделили минимум два эксперта.

Ниже мы опишем все этапы разработки и тестирования системы, извлекающей эксплицитную оценку, а затем дополненной модулем извлечения имплицитной оценки, и сравним полученные результаты.

2.2. Построение и тестирование системы извлечения явной оценки

2.2.1 Составление оценочного словаря

Словарь оценочных слов был составлен вручную. Для его составления использовались следующие источники:

1. Список оценочных слов, доступный на сайте РОМИП (romip.ru), автоматизировано сформированный авторами работы [Chetviorkin, Loukachevitch; 2012]; данный список содержит слова, извлеченные из коллекции отзывов и не разделенные по тональности. Для использования в настоящей работе слова из списка были экспертно размечены на позитивные и негативные, а затем автоматически разбиты на группы по частям речи.
2. «Словарь русской идиоматики. Сочетания слов со значением высокой степени» Г.И. Кустовой. Из данного словаря также были вручную выделены положительные и отрицательные оценочные слова, а также слова (преимущественно существительные), носящие позитивную или негативную коннотацию.

3. Обработанный вручную корпус политических блогов (200000 слов), в котором все слова, выражающие оценку, размечены по тональности и части речи. Внесение в словарь тональной лексики из данного корпуса, по тематике аналогичного тестовому, позволяет сделать конструируемую систему более предметно-ориентированной, тем самым увеличив точность ее работы.

В результате совмещения полученных списков был сформирован словарь объемом 2700 слов и словосочетаний (к примеру, *сойти с ума*), слова в котором впоследствии были разделены на группы в соответствии с их семантическими и синтаксическими особенностями для последующего подставления в шаблоны. Классификация лексики и присвоенные группам обозначения представлены ниже.

1. Позитивно-оценочные слова

- 1.1.1 Положительные прилагательные и наречия – *pos_adj* и *pos_adv*;;
 - 1.2 Положительные существительные
 - 1.2.1 Положительно-оценочные существительные (*молодец, герой*) – *pos_est_noun*;
 - 1.2.2 Положительные качества человека (*доброта, мудрость*) – *pos_trait_noun*;
 - 1.2.3 Существительные, имеющие положительное значение только в сочетании с нейтральным или положительным контекстом (*рост*) – *potential_pos_noun*;
 - 1.2.4 Существительные положительного отношения (*уважение, благодарность*) – *pos_attitude_noun*;
 - 1.3 Положительные глаголы
 - 1.3.1 Положительно-оценочные глаголы – предикаты субъекта (автора) высказывания (*любить, обожать*) – *pos_subj_verb*;
 - 1.3.2 Положительно-оценочные глаголы– предикаты объекта высказывания (*нравиться, радовать*) – *pos_obj_verb*;
 - 1.3.3 Глаголы – предикаты объекта высказывания, носящие положительную коннотацию (*улучшать*) – *pos_con_verb*;

2. Негативно-оценочные слова

- 2.1 Отрицательные прилагательные и наречия – *neg_adj*, *neg_adv*;
- 2.2 Отрицательные существительные

2.2.1 Отрицательно-оценочные существительные (*идиом*) – *neg_est_noun*;

2.2.2 Отрицательные качества человека (*идиотизм*) – *neg_trait_noun*;

2.2.3 Существительные, имеющие отрицательную коннотацию (*кровь, ненависть*) – *neg_con_noun*;

2.2.4 Существительные, имеющие отрицательное значение только в сочетании с нейтральным или положительным контекстом (*падение*) – *potential_neg_noun*;

2.2.5 Существительные отрицательного отношения (*брезгливость*) – *neg_attitude_noun*;

2.3 Отрицательные глаголы

2.3.1 Отрицательно-оценочные глаголы – предикаты субъекта (автора) высказывания (*ненавидеть*) – *neg_subj_verb*;

2.3.2 Отрицательно-оценочные глаголы – предикаты объекта высказывания (*бесить*) – *neg_obj_verb*;

2.3.3 Глаголы – предикаты объекта высказывания, носящие отрицательную коннотацию (*ухудшать*) – *neg_con_verb*;

Что касается глаголов, семантика которых зависит от тональности контекста (например, «воевать с преступностью»/ «воевать с народом»), то содержащие их конструкции мы будем относить к имплицитной оценке (как переносящие и инвертирующие тональность другого объекта) и рассмотрим при создании модуля ее извлечения.

Следует заметить, что приведенные списки могут пересекаться – к примеру, списки положительных и отрицательных качеств являются подмножествами списков существительных с положительной и отрицательной коннотацией.

Списки прилагательных и наречий объединены из-за особенностей работы используемого инструмента (*tomita-parser*), использующего одновременно все разборы слова; таким образом, все наречия, образованные от прилагательных, могут быть найдены по форме прилагательного из-за омонимичности форм наречия и краткого прилагательного среднего рода. Наречия, не образованные от прилагательного, добавлены к списку отдельно.

2.2.2 Составление списков слов, участвующих в выражении оценки

Помимо оценочных слов, в выделении оценки также участвуют и другие классы лексики. Для составления правил, аналогичных описанным в [Ермаков, Киселев; 2005] и

[Пазельская, Соловьев; 2011] мы составили требуемые списки лексем, частично позаимствовав их из указанных работ. К ним относятся:

- Инверторы (не, отсутствие, ликвидация) – invert;
- Глаголы отождествления (являться, представлять собой) equate_verb; кроме того, сюда же мы относим конструкции «стать/оказаться Y-ом», так как они выражают оценку того же (хоть и изменившегося в глазах субъекта высказывания) объекта;
- Прилагательные-интенсификаторы, обозначающие низкую степень (маленький, низкий) – lower_adj и высокую – (высокий, огромный) – heighten_adj;
- Наречия недостаточной степени (мало, недостаточно) – not_enough_adv и избыточной степени (слишком, чересчур) – excessing_adv. При этом наречия недостаточной степени в нашем словаре являются подклассом негативных наречий; это связано с тем, что в сочетании с положительно-оценочными прилагательными ранее негативные наречия приобретают значения интенсификации, но сохраняют негативную окраску в сочетании с глаголом («ужасно красив» vs «ужасно пишет»), в то время как наречия недостаточной степени придают как признаку, так и действию отрицательную полярность («недостаточно красивый», «недостаточно работает»). Что касается наречий избыточной степени, то они сообщают негативную тональность только нейтральным прилагательным, не меняя полярность оценочных («слишком красивый»). Сочетание таких наречий с нейтральными глаголами нами учитываться не будет, так как, на наш взгляд, не носит достаточной негативной окраски («он слишком торопится»).
- Маркеры неверидикативности []:
 - Вопросительный знак в конце предложения
 - «Если» перед тональной пропозицией в сочетании с глаголом будущего времени
 - «Бы» перед тональной пропозицией
 - «Якобы» перед тональной пропозицией
 - «Чтобы» перед тональной пропозицией
 - «Пока не» перед тональной пропозицией
 - «Сомневаюсь, что» перед тональной пропозицией
 - «Маловероятно, что»
 - «invert уверен/думаю/считаю, что» перед тональной пропозицией (*«едва ли уверен, что»*)

- Повелительное наклонение глаголов, обладающих положительной/отрицательной коннотацией.
- Сравнительная степень любых прилагательных и наречий

2.2.3 Построение правил извлечения оценки

Релевантный объект (то есть объект, оценка которого нас интересует) обозначен как *rel_obj*; в качестве этого объекта может выступать именная группа (единицы которой не входят в оценочные словари), персона, организация, личное местоимение и местоимение «который». В Таблице 1 представлены правила выделения релевантного объекта, написанные на языке tomita-парсера.

Таблица 1. Правила выделения релевантного объекта.

Rel_ob → SPRO "который";	Личное местоимение
Rel_obj → Word <kwset="должность"> Word<kwset="фио">;	(президент) Дмитрий Медведев -персона, с должностью (из словаря должностей) или без; <i>фио</i> – встроенный алгоритм выделения персоны в любой форме (Пушкин, А.С. Пушкин, Александр Сергеевич и т.д.)
Rel_obj → Word <kwset="форма организации"> Organization ; Rel_obj → Organization ; Organization → ABBR+ Word<gram="geo"> Org_noun_phrase ABBR Org_noun_phrase Word<h-reg1>; ABBR → Word<wfm="[А-Я][А-Я]+"> ; Org_noun_phrase → AdjP_first<c-agr[1]> (AdjP_second<c-agr[1]>)* Noun<c-agr[1], kwset=~"словарь_оценочных_слов"> Noun_phrase_gen* PP; Org_noun_phrase → AdjP_first<c-agr[1]> (AdjP_second <c-agr[1]>)* Noun <c-agr[1],	ООО «РПП», ОАО РПП, ООО Ромашка, компания Газпром, Московский государственный университет, Россия, IX съезд полярников, Комитет по безопасности (искусственное ограничение на одну предложную группу с предлогом «по»)

<p>kwset=~"словарь_оценочных_слов"> Noun_phrase_gen*; Org_noun_phrase → Noun<gram="S", h-reg1, kwset=~"словарь_оценочных_слов"> Noun_phrase_gen* PP; Org_noun_phrase → Noun<gram="S", h-reg1, kwset=~"словарь_оценочных_слов"> Noun_phrase_gen*; AdjP_first → Word<gram="A", h-reg1, kwset=~"словарь_оценочных_слов"> Word<wfm="[XIV]+"> Word<gram="NUM", h- reg1>; AdjP_second → Word<gram="A", kwset=~"словарь_оценочных_слов">; PP → "по" Noun <kwset=~"словарь_оценочных_слов"> Noun_phrase_gen *; Noun_phrase_gen → Word<gram="A, gen", kwset=~"словарь_оценочных_слов">* Noun<gram="S,gen", kwset=~"словарь_оценочных_слов"> Noun_phrase_gen*;</p>	
<p>Rel_obj→NP; NP→AdjP<c-agr[1]> Noun <c-agr[1], kwset=~"словарь_оценочных_слов">; NP→AdjP<c-agr[1]> Noun <c-agr[1], kwset=~"словарь_оценочных_слов"> Noun_phrase_gen; AdjP→ADV* Word<gram="A", kwset=~"словарь_оценочных_слов"> APRO;</p>	<p>Именная группа в любом падеже, ни одна единица из которой не входит в словарь оценочной лексики.</p>

Noun_phrase_gen → Word<gram="A, gen", kwset=~"словарь_оценочных_слов">* Noun<gram="S,gen", kwset=~"словарь_оценочных_слов"> Noun_phrase_gen*;	
---	--

Далее представим в упрощенном виде (на псевдо языке продукционных правил, аналогичных правилам tomita-парсера) набор написанных нами правил извлечения оценки, используя в качестве терминалов введенные обозначения групп оценочной лексики и общепринятые обозначения частей речи.

Все правила проверялись на Национальном корпусе русского языка (ruscorpora.ru).

Таблица 2. Правила выделения положительной оценки

pos_adj_phrase → ^invert adv (^not_enough_adv)* ^invert pos_adj	«симпатичный», «потрясающе красивый», «страшно симпатичный», и даже «слишком красивый», но не «недостаточно красивый» – положительное прилагательное без инверсии, с наречием или без; при этом отсутствием инверсии будет считаться нулевое или четное ее количество
pos_adj_phrase → adv* invert neg_adj	«совсем не неказистый» - инверированное отрицательное прилагательное с наречием или без;
pos_adj_phrase → invert (очень/слишком) neg_adj	«не слишком плохой»

neutral_adj_phrase → adv <kwset= ~"словарь_оценочных_слов"> adj <kwset= ~"словарь_оценочных_слов">	Нейтральная группы прилагательного
neutral_noun_phrase → adv <kwset= ~"словарь_оценочных_слов"> noun <kwset= ~"словарь_оценочных_слов">	Нейтральная именная группа
positive_noun → (^invert) pos_est_noun (^invert) pos_con_noun invert neg noun	положительное существительное – это положительно-оценочное существительное без инверсии, существительное с положительной коннотацией без инверсии или отрицательное существительное с инверсией; отсутствием инверсии считается нулевое или четное ее количество.
pos_noun_phrase → pos_adj_phrase pos-noun pos_adj_phrase noun <kwset= ~"словарь_оценочных_слов">	« <i>деятельный руководитель</i> », но не « <i>деятельный дурак</i> »
pos_noun_phrase → neutral_adj_phrase* positive_noun	« <i>известный герой</i> », « <i>гуманизация</i> »
pos_noun_phrase → potential_pos_noun (neutral_noun_phrase neutral_adj_phrase pos_ noun)	« <i>рост промышленности</i> », « <i>рост гуманизации</i> », но не « <i>рост преступности</i> »
pos_noun_phrase → potential_neg_noun уровня* neg_noun_phrase	« <i>падение (уровня) преступности</i> »
pos_noun_phrase_1 → (^invert) neutral_noun_phrase neutral_noun_phrase(gen)* pos_noun_phrase (gen)	« <i>пример мудрости</i> », « <i>выражение добродетели</i> » – цепочка существительных в родительном падеже является положительной, если, во-первых, главное слово положительное и, во-вторых, в числе зависимых слов нет отрицательных (в том числе и потенциально отрицательных) и инверторов
pos_noun_phrase → noun(^neg) без neg_noun (gen)	« <i>мир без преступности</i> »
pos_proposition → rel_obj – (это)	« <i>Петров – хороший человек</i> », « <i>Петров –</i>

pos_noun_phrase;	<i>герой», «обещание выйти из договора – достойная реакция»; «коммунизм – это рост промышленности»</i>
pos_proposition → rel_obj equate_verb pos_noun_phrase	<i>«Петров является хорошим руководителем»</i>
pos_proposition → «считать» (1 л.) rel_obj (Gen) pos_noun_phrase (Instr)	<i>«считаю Петрова хорошим руководителем»</i>
pos_proposition → pos_adj rel_obj	<i>«чудесная книга»</i>
pos_proposition → pos_trait_noun rel_obj (Gen);	<i>«доброта Петрова»</i>
pos_proposition → (меня нас мне нам) pos_obj_verb rel_obj;	<i>«мне нравится эта книга»</i>
pos_proposition → (^Noun (gen) ^ему ^ей ^его ^ее) pos_obj_verb rel_obj;	<i>«радует книга Петрова», но не «ей нравится эта книга»</i>
pos_proposition → rel_obj pos_con_verb;	<i>«президент улучшил состояние промышленности»</i>
pos_proposition → относиться (1л.) с (adj) pos_attitude_noun к rel_obj (Dat) pos_proposition → относиться (1л.) с (adj) pos_attitude_noun к таким neutral_noun_phrase (plural, dat), как rel_obj	<i>«я отношусь с глубочайшим уважением к Петрову» «отношусь с уважением к таким людям, как Петров»</i>
pos_proposition → выражать (1л.) (adj) pos_attitude_noun rel_obj (Dat);	<i>«выражаю глубокое уважение Петрову»</i>
pos_adv_phrase → ^neg_adv pos_adv	<i>«хорошо», но не «недостаточно хорошо»</i>
pos_proposition → rel_obj ^invert pos_adv_phrase verb<kwset= ~"словарь_оценочных_слов">	<i>«Иванов хорошо пишет»</i>
pos_proposition → такой pos_noun_phrase, как rel_obj	<i>«такой хороший человек, как Иванов»</i>
pos_proposition → rel_obj – один из pos_noun_phrase (gen)	<i>«Петров – один из лучших журналистов»</i>

pos_proposition → pos_noun_phrase rel_obj (person)	«хороший журналист Петров»
pos_proposition → rel_obj – человек (heighten_adj (gen) neutral_noun_phrase) (pos_noun_phrase)(^высокого роста)	«Петров – человек высокой честности/незаурядного интеллекта/огромного потенциала/большой доброты», но не «высокого роста»
pos_proposition → rel_obj – человек с pos_noun_phrase	«Петров – человек с отличными качествами/хорошим чувством юмора»

Таблица 3. Правила выделения отрицательной оценки

negative_adj_phrase → (^invert) adv (^not_enough) neg_adj invert pos_adj not_enough_adv pos_adj (not_enough_adv excessing_adv) adj <kwset= ~"словарь_оценочных_слов">	«плохой», с наречием или без (кроме «недостаточно плохой»), «не красивый», «недостаточно красивый», «недостаточно длинный», «слишком длинный»; при этом отсутствием инверсии считается нулевое или четное ее количество.
negative_noun → (^invert) neg_est_noun (^invert) neg_con_noun invert pos_noun ” pos_noun”;	отрицательное существительное – это отрицательно-оценочное существительное без инверсии, существительное с негативной коннотацией без инверсии или положительное существительное с инверсией, или положительное существительное в кавычках; отсутствием инверсии считается нулевое или четное ее количество)

neg_noun_phrase → negative_adj_phrase noun;	«плохой руководитель», «плохой защитник», «злостный нарушитель»
adj_phrase → adv adj neg_noun_phrase → adj_phrase neg_noun;	«потрясающий негодяй», «совершенно неинтересная нудятина»
neg_proposition → rel_obj – (это) neg_noun_phrase;	«Петров – плохой человек», «Петров – дурак», «современные выборы – это профанация»; «коммунизм – рост промышленности»
neg_assessment → rel_obj equate_verb neg_noun_phrase	«Петров является плохим руководителем»
neg_assessment → opinion_verb (1 л.) rel_obj (Gen) neg_noun_phrase (Instr);	«считаю Петрова хорошим руководителем»
neg_assessment → neg_adj rel_obj;	«мерзкая книга»
neg_assessment → neg_trait_noun rel_obj(Gen);	«глупость Петрова»
neg_assessment → neg_est_noun rel_obj(Gen);	«прихвостни Петрова»
neg_assessment → (меня нас мне нам) neg_obj_verb rel_obj;	«меня раздражает эта книга»
neg_assessment → (меня нас мне нам) invert pos_obj_verb rel_obj;	«мне не нравится эта книга»
neg_assessment → (^Noun (gen) ^ему ^ей ^его ^ее) neg_obj_verb rel_obj	«раздражает книга Петрова»
neg_pronegition → относиться (1л.) с (adj) neg_attitude_noun к rel_obj (Dat) neg_pronegition → относиться (1л.) с (adj) neg_attitude_noun к таким neutral_noun_phrase	«с брезгливостью отношусь к таким людям» «с брезгливостью отношусь к таким людям, как Петров»

(plural, dat), как rel_obj	
neg_pronegation → neg_noun_phrase rel_obj (person)	«плохой руководитель Петров»
pos_proposition → такой pos_noun_phrase, как rel_obj	«такой плохой руководитель, как Петров»
pos_proposition → rel_obj – один из pos_noun_phrase (gen)	«Петров – один из тех ужасных людей...»
neg_assessment → rel_obj adv* neg_con_verb;	«президент ухудшил состояние промышленности» (недостаточно ухудшил), «потрясающе ухудшил»)
neg_proposition → rel_obj – человек (lower_adj (gen) neutral_noun_phrase) (neg_noun_phrase (gen)) (^низкого роста)	«Петров – человек низкого развития» (но не «низкого роста») «Петров – человек ужасного воспитания»
neg_pronegation → rel_obj – человек с neg_noun_phrase	«Петров – человек с ужасным характером»

Данные правила представлены в обобщённом виде; в продукционном классификаторе, написанном с помощью tomita-парсера, присутствуют вариации данных правил, учитывающие свободный порядок слов, наличие посторонних слов («также», вводные слова и т.д.) и сочинительных групп.

2.2.4 Тестирование системы

Тестовый корпус состоит из текстов блогов политической направленности и содержит 100 записей (около 50000 слов).

Корпус размечался параллельно тремя экспертами. Инструкция разметки состояла из нескольких правил:

1. Выделить все объекты, отношение автора к которым они могут охарактеризовать как положительное и отрицательное;
2. Приписать выделенным объектам тональность;
3. Объектом может являться именная группа, имя собственное или личное местоимение, а также местоимение «который»;

4. Выделять конкретный объект в рамках одного предложения, где была высказана оценка, игнорируя анафору и кореференцию (т.е. во фразе «Также в деле выступал Петров. Прекрасный, добрейший человек» объектом оценки будет только «человек»).

Для результирующей разметки были выбраны те объекты, которые выделили минимум два эксперта (что касается тональности оценки, то в ее разметке значительных расхождений не наблюдалось). Таким образом было оставлено 270 объектов.

Результаты работы системы оценивались по двум параметрам: правильное выделение оцененного объекта и правильное выделение тональности оценки выделенных объектов (возможна только оценка точности). Результаты представлены ниже.

	Preciseness	Recall	F-measure
Выделение объекта оценки	0,92	0,3	0,46
Определение полярности оценки	0,863		

2.3. Построение модуля извлечения скрытой (опосредованной) оценки

2.3.1 Выявление правил извлечения скрытой оценки

В размеченном нами обучающем корпусе было выделено четыре типа конструкций-отношений между объектами: переносящие позитивную оценку, переносящие негативную оценку, инвертирующие позитивную и негативную оценку. В общей сложности было выделено 300 примеров, которые впоследствии были сгруппированы в классы в соответствии с семантикой входящих в них элементов для последующего расширения этих классов.

Участники описанных ситуаций – это X (релевантный объект) и Y (явно оцененный в тексте или имеющий положительную/отрицательную коннотацию объект, который может быть именной группой или глаголом в неопределенной форме («ему нравится издеваться»)). Классы выделенных конструкций приведены ниже.

1) X хорошо относится к Y-у/ X помогает Y-у

К данному классу относятся конструкции, в которых явно оцененный объект в определенном смысле выступает бенефактивом, получая какую-либо выгоду от действия X (релевантного объекта).

Этот класс довольно ограничен, поэтому мы экспертно расширили его при помощи синонимичных конструкций из **словаря синонимов**. Все добавленные конструкции иллюстрируются примерами – сконструированными или найденными в Национальном корпусе русского языка. Кроме того, уже при составлении правил для tomita-парсера мы учитывали замену глаголов отглагольными существительными (X поддерживает Y = поддержка Y-а X-ом).

Добавлено примечание ([AC1]): словарь

Таблица 4. Перенос оценки через пассивную поддержку оцененного объекта.

Шаблон	Пример
X любит Y	<i>«президент любит таких серых выскочек»</i>
X-у нравятся Y	<i>«президенту нравятся такие серые выскочки»</i> (сконструированный пример)
X одобряет Y	<i>«члены ЕС одобряют и поощряют бомбардировки и уничтожение Сирии как независимого государства»</i>
X не осуждает Y	<i>«ЕС не осуждает неправомерные действия США»</i>
X поощряет Y	<i>«члены ЕС одобряют и поощряют бомбардировки и уничтожение Сирии как независимого государства»</i>
X согласен с Y-ом	<i>«Пономарев с неправосудным решением Нью-Йоркской Фемиды полностью согласен»</i>
X (находится/состоит) в pos_adj отношениях с Y-ом	<i>«местные власти в прекрасных отношениях к криминальными авторитетами»</i>
X с Y-ом – друзья	<i>«Либералы с жестокими ментами – друзья-братья»</i>

X - сторонник Y	<i>«Эти христианские лица – сторонники скандального законопроекта»</i>
X на стороне Y	<i>«они на стороне скандального законопроекта»</i> (сконструированный пример)

Как видно из приведенных примеров, в размеченном нами корпусе не встретилось ни одного случая переноса положительной оценки вышеперечисленными глаголами; действительно, положительное отношение объекта к чему-то хорошему не переносит оценку на сам объект (*«ему нравятся хорошие книги»*). Таким образом, можно сделать вывод, что глаголы данного подкласса могут переносить только негативную оценку.

Следует также заметить, что такой перенос оценки работает только в одну сторону, с объекта действия на субъект действия, но не наоборот, о чем свидетельствует пример, приведенный в [Ермаков, Киселев; 2005]: *«олигархи любят президента»* (да мало ли за что?).

Таблица 5. Перенос оценки через активную поддержку оцененного объекта.

X заботится о Y -е	<i>«сколько наших набежало, заботящихся о наших коррупционерах»</i>
X помогает Y-у	<i>«правительство помогает негодьям»</i> <i>«институт помогает нам защищаться от недобросовестных политиков»</i> <i>«эта мера поможет уменьшить преступность и защитить людей»</i>
X выступает в защиту Y	<i>«суды и прокуратура выступают в защиту негодяев в погонах»</i>
X защищает Y	<i>«суды и прокуратура защищают негодяев в погонах»</i> (сконструированный диплом)

Х – защитник Y-а/ защитник Y-а X	<i>«опозиционеры – защитники крамольных девок»</i>
поддержка X-ом Y-а	<i>«Поддержка болотными активистами акции осквернения Храма Христа Спасителя ...»</i>
X поддерживает Y	<i>«Правительство поддержало сомнительный законопроект»</i>
X финансирует Y	<i>«те, кто финансирует преступников»</i>
X спонсирует Y	<i>«те, кто спонсирует преступников»</i> (сконструированный пример)
X – спонсор Y (X <i>equate_verb</i> Y)	<i>«те, кто стал спонсорами преступников»</i> (сконструированный пример) <i>«Кремль – гнездовье спонсоров мирового терроризма»</i>
X кормит Y	<i>«кормящие недобросовестных чиновников бизнесмены»</i>

Как видно из таблицы, примеров переноса оценки с положительно оцененного объекта на релевантный почти не встретилось. С другой стороны, было найдено немало примеров, где данные глаголы, сочетаясь с нейтральным контекстом, положительно характеризовали релевантный объект (субъект действия) – например, в примере *«Pussy Riot защищают естественное право человека»* ни одно слово из именной группы, выражающей объект действия, не присутствует в словаре оценочной лексики. Аналогично, в работе [Пазельская, Соловьев; 2011] многие из перечисленных глаголов принадлежат классу «потенциально положительных», то есть положительно характеризуют субъект действия в сочетании с положительным или нейтральным объектом действия. Мы также согласимся с этой концепцией, касающейся всех перечисленных глаголов и словосочетаний, за исключением выражений «спонсирования», которые сами по себе не носят положительной коннотации.

Такие выражения интересны и с другой точки зрения: в отличие от остальных конструкций данного класса, они могут передавать оценку и в обратном направлении, что

следует из встретившегося в корпусе примера *«среди спонсоров Обамы называют криминальный бизнес»*.

2) Y – содержание действия X

Конструкции данного класса также обобщены в соответствии с ролями входящих в них единиц, когда оцененный объект Y является содержанием действия агенса X. Такое описание ролей совпадает с их представлением в системе Framebank (framebank.ru), что позволяет расширить данный класс, добавив в список ряд конструкций с аналогичными фреймами.

Таблица 6. Перенос оценку через оценку деятельности релевантного объекта

Шаблон	Пример
X зарабатывает на Y-е	<i>«Современные нувориши, зарабатывающие на сомнительных и низкопробных шоу»</i>
X занимается Yом	<i>«"цапки" занимались не только убийствами и вымогательствами» «учителя занимаются антиконституционной деятельностью»</i>
X делает Y	<i>«они делают ужасные вещи» «Павел Пряников сделал отличную подборку» «(ты) начала делать серьезные вещи»</i>
X совершает Y	<i>«государство Российская Федерация, совершив беззаконие...»</i>
X творит Y	<i>«руководство РПЦ понимает, что творит зло»</i>
X упражняется в Y	<i>«пока Евросоюз упражняется в лицемерии...»</i>
X совершенствуется в Y	<i>«спецслужбы совершенствуются в противозаконных методах контроля»</i>
X – чемпион (мира) по Y-у	<i>«марксисты чемпионы мира по ...расстрелам, вранью и сокрытию»</i>

	<i>преступлений»</i>
X ответил Y-ом	<i>«он ответил неадекватными действиями»</i>

Основываясь на данных системы Framebank, мы расширили данный класс такими фреймами, как «начинать acc/inf», «заканчивать acc/inf», «решить inf», «суметь inf», «готовиться к dat», «вести acc», «заклучаться в rger», «заслужить acc», «исполнять acc», «играть acc», «ограничиваться ins», «оказать acc», «отплатить ins», «перейти к dat», «поспешить inf», «потрудиться inf», «поторопиться inf», «привыкнуть inf», «применить acc», «приступить к dat», «собраться inf», «способствовать inf», «успеть inf», «хотеть inf».

3) Y – результат деятельности X

В конструкциях данного класса Y является результатом действия агенса X (а иногда и самого существования X, если последний обозначает явление или процесс). Такое описание ролей также совпадает с их представлением в системе Framebank и представляет обширный класс глаголов; извлеченный из Framebank список представлен в Приложении 1, а начальный список с примерами из размеченного корпуса – в следующей таблице.

Таблица 7. Оценка объекта через оценку результата его деятельности.

Шаблон	Пример
X строит Y	<i>«Американский президент строит свою диктатуру»</i>
X устраивает Y	<i>«Евреи и Палестинцы опять устроили мордобой с ракетами»</i>
X изготавливает Y	<i>«Тухлое политическое блюдо, изготовленное властью»</i>
X – автор Y	<i>«Сидякин – автор подленького закона...»</i>
X производит Y	<i>«она производила антидемократические законы ежедневно»</i> <i>«ты очень хорошее впечатление на всех произвел»</i>

	<i>«они производят замечательный продукт»</i>
X готовит Y	<i>«известным радикалом ..., готовившим ряд крупных терактов»</i>
X предлагает Y	<i>«они голосовали за человека, который предлагал НЕ конструктивные методы решения»</i>
X устанавливает Y	<i>«путин давно установил антидемократический режим»</i>
X принимает Y	<i>«путинисты приняли пакет репрессивных законов»</i>
X проводит Y	<i>«и ведь сколько акций отвратительных они провели» «Урлашов провел честные выборы»</i>
X создает Y	<i>«они создали незаконное государство» «Дмитрий Солопов создал отличную радиостанцию»</i>
X организует Y	<i>«члены Евросоюза организовали бомбардировки Ливии»; «спецслужбы, организующие фальшивые сливы»</i>
X посеял Y	<i>«Вы... посеяли апатию»</i>
X порождает Y	<i>«...сочинение всяческих текстов, которые порождают потом споры и перепалки»</i>
X ведет (Z) к Y-у	<i>«тандем очевидно ведет страну к развалу»</i>
X привел к	<i>«он привел к развалу государственных институтов»</i>
X способствует Y-у	<i>«закон способствует росту коррупции»</i>
X благодатная почва для Y-а	<i>«Это благодатная почва для мелких жуликов»</i>
X закончился Y-ом	<i>«попытка ... закончится»</i>

	<i>тотальным позором»</i>
X кончится Y-ом	<i>«попытка... кончится насилием и революцией»</i>
est_verb/Y по приказу X-а	<i>«сожженных по его приказу Бесланских детей»</i>
est_verb/Y по указанию X-а	<i>«по указанию нач. ГУВД была грубо пресечена попытка проведения флеш-моба»</i>
est_verb/Y по воле X-а	<i>«журналиста могли избить или убить по воле обнаглевшего чиновника»</i>
X дает/отдает приказ/распоряжение est_verb	<i>«генералы дают приказы расправиться с журналистами»</i>

Говоря о двух последних типах конструкций переноса оценки, следует заметить, что такие отношения между субъектом и объектом действия могут переносить как отрицательную, так и положительную оценку (большее число примеров с отрицательной оценкой, на наш взгляд, свидетельствует только об ее большей распространённости); при этом оценка переносится с объекта действия на его субъект, но не наоборот («этот ужасный автор написал новую книгу» - не факт, что книга ужасная).

4) Тесная связь между X и Y

В данный класс объединены, по сути, несколько типов конструкций – непосредственное выражение тесной связи между релевантным и оцененным объектами, отношения типа часть-целое, а также иерархические отношения подчинения. Особенностью таких связей является их двунаправленность, обусловленная близким отношением объектов, иногда приводящим к их отождествлению. Этот класс конструкций также был расширен ближайшими синонимами.

Таблица 8. Оценка объекта через тесно связанный с ним объект.

Тесные взаимоотношения X с Y	<i>«Тесные взаимоотношения Патриарха с бандитами»</i>
Крепкая связь между X и Y	<i>«Крепкая связь между ней и криминальным бизнесом»</i>

X крепко/тесно связан с Y	<i>«он тесно связан с криминальным бизнесом» (сконструированный пример)</i>
X в связке с Y	<i>«рейдеры (т.е. бандиты), которые в связке с судебными... органами...»</i>
За X-ом (стоит) Y	<i>«За компанией «Рубикон» стоят обычные рейдеры (т.е. бандиты)» «за коммунизмом дикость и зверства...»</i>
X подконтролен Y-у	<i>«правоохранительные органы давно подконтрольны жуликам и ворам»</i>
X подчиняется Y-у	<i>«жизнь российской бюрократии стала полностью подчиняться мафиозным принципам»</i>
X строится на Y-е	<i>«кампания ... строится на откровенной лжи и подтасовке фактов»</i>
Y характерен для X	<i>«такие игрища характерны для современного православного сознания»</i>
Черта X – его Y Чертой X equate_verb ergo Y	<i>«Отличительной чертой Никифорова как прокурора... была его профессиональная непригодность и профессиональная беспомощность» «Портал, отличительная черта которого – проверенная, достоверная информация»</i>
X демонстрирует (свой) Y	<i>«Кремль и федералы постоянно демонстрируют обществу манипуляцию и пренебрежение законами» «Она невероятно активная, демонстрирующая незаурядные организаторские способности...»</i>
X показывает (свой) Y	<i>«Кремль и федералы постоянно показывают пренебрежение законами» (сконструированный пример)</i>
X известен (своим) Y-ом	<i>«господин Либерман, известный</i>

	<i>своими фашистскими взглядами и кривым акцентом»</i>
Y со стороны X	<i>«подвергся несправедливым преследованиям со стороны местных властей»</i>
X equate_verb инструментом Y-а	<i>«суд является инструментом расправы»</i>

Данные конструкции могут переносить и отрицательную, и положительную оценку, но, как мы видим, к переносу положительной оценки наиболее склонны конструкции, в которых Y выступает как черта/характер объекта.

5) X против Y; отношения, инвертирующие оценку.

Таблица 9. Оценка объекта через инвертируемую оценку другого объекта.

X противоречит Y	<i>«Восстановление этой статьи в УК противоречит принципам демократизации общества»</i>
X выступает против Y	<i>«Парень, выступающий против ненависти к другим народам» «Pussy Riot выступили против авторитарного режима»</i>
X борется с Y-ом	<i>«президент борется с коррупцией»</i>
X воюет с Y-ом X объявил войну Y-у	<i>«Си Цзиньпин объявил войну партийной коррупции»</i>
X осудил/не поддержал Y	<i>«Правительство не поддержало сомнительный законопроект» «Полно мусульман, готовых осудить террор»</i>
X исключает (возможность) Y	<i>«Отсутствие общекавказской</i>

	<i>консолидации правозащитных прав исключает возможность содержательной и системной организованной работы»</i>
--	--

Конструкции данного типа могут инвертировать оценку любой полярности.

Кроме того, некоторые глаголы данного класса являются потенциально-отрицательными ([Пазельская, Соловьев; 2011]).

2.3.2 Тестирование системы, дополненной модулем извлечения скрытой оценки.

Таким образом, мы сформулировали полученные правила на языке tomia-парсера (учитывая возможные вариации порядка слов) и добавили их к правилам выделения явной оценки, а затем протестировали на том же корпусе, что и первый вариант системы. Результаты тестирования представлены ниже.

	Preciseness	Recall	F-measure
Выделение объекта оценки	0,87	0,422	0,56
Определение полярности оценки	0,85		

2.4. Выводы

Итак, мы протестировали созданный нами макет системы извлечения явной оценки, построенной на продукционных правилах, а затем дополнили его модулем извлечения скрытой оценки и сравнили результаты. Оценка результатов показала увеличение F-меры на 10% за счет значительного увеличения полноты и незначительного снижения точности, что позволяет сделать вывод о целесообразности использования добавленных правил.

Кроме того, была незначительно снижена точность определения полярности оценки; по результатам анализа ошибок такого рода выяснено, что чаще всего они обусловлены ошибочным подсчетом количества инверторов в правиле или отсутствием учета в правилах некоторых видов ирреального контекста.

Заключение

Таким образом, в рамках настоящего исследования нами были решены следующие задачи:

1. Создание и тестирование макета системы извлечения явной оценки объектов, основанной на продукционных правилах.
В процессе решения данной задачи, во-первых, был составлен словарь оценочных слов и выражений и, во-вторых, были исследованы и уточнены правила извлечения явной оценки.
2. Выделение и классификация шаблонов переноса оценки на материале корпуса политических блогов, а также расширение данных классов при помощи словаря синонимов и фреймов семантических ролей системы FrameNet.
Анализ выделенных конструкций показал, что отрицательная оценка способна распространяться на другие объекты не только чаще положительной, но и с использованием более широкого арсенала конструкций.
3. Тестирование системы, дополненной модулем извлечения скрытой оценки и сравнение результатов.
Сравнительный анализ результатов работы системы показал, что добавление модуля извлечения скрытой оценки дает значительное увеличение качества

работы системы и подтверждает целесообразность его использования (в первую очередь, на материале текстов политической направленности).

В качестве дальнейших исследований по данной тематике можно выделить следующие направления:

1. Создание полноценной системы извлечения явной и скрытой оценки, имеющей модуль разрешения анафорических и кореферентных связей;
2. Уточнение правил выделения оценки и повышение качества их работы;
3. Исследование явления скрытой оценки на других типах текстов.

Список литературы

1. Andreevskaia Alina And Sabine Bergler. 2006a. Mining Word-Net For Fuzzy Sentiment: Sentiment Tag Extraction From Wordnet Glosses. // In Proceedings Of The 11th Conference Of The European Chapter Of The Association For Computational Linguistics (Eacl'06), Pages 209–216, Trento, It. [Http://Www.Aclweb.Org/Anthology-New/E/E06/E06-1027.Pdf](http://www.aclweb.org/anthology-new/E/E06/E06-1027.pdf)
2. Andreevskaia Alina, Nancy Ide. 2007. Tutorial: Overview: Sentiment And Opinion Analysis. // Eurolan 2007 Lecturers. Slides. [Http://Eurolan.Info.Uaic.Ro/Html/Profs/Materials/Ide.Pdf](http://eurolan.info.uaic.ro/html/profs/materials/ide.pdf)
3. Andreevskaia, A. And S. Bergler: 2007, 'Clac And Clac-Nb: Knowledgebased And Corpus-Based Approaches To Sentiment Tagging'. // In: 4th International Workshop On Semantic Evaluations (Semeval 2007). Prague, Czech Republic. [Http://Acl.Ldc.Upenn.Edu/W/W07/W07-2022.Pdf](http://acl.ldc.upenn.edu/W/W07/W07-2022.pdf)
4. Aue, A. And M. Gamon: 2005, 'Customizing Sentiment Classifiers To New Domains: A Case Study'.// In: Ranlp-05, The International Conference On Recent Advances In Natural Language Processing. Borovets, Bulgaria. [Http://Research.Microsoft.Com/~Anthaue/New_Domain_Sentiment.Pdf](http://research.microsoft.com/~antheaue/new_domain_sentiment.pdf)
5. Beineke, P.; Hastie, T.; Vaithyanathan, 2004 The Sentimental Factor: Improving Review Classification Via Human-Provided Information. // In Proceedings Of Acl-04,

- The 42nd Annual Meeting Of The Association For Computational Linguistics, Barcelona, Es, 2004, Pp. 263–270.
6. Bethard, S., H. Yu, A. Thornton, V. Hatzivassiloglou, And D. Jurafsky, D. 2004. Automatic Extraction Of Opinion Propositions And Their Holders. // In James G. Shanahan, Janyce Wiebe, And Yan Qu, Editors, Proceedings Of The Aaai Spring
 7. Chetviorkin I., Loukachevitch N. Evaluating sentiment analysis systems in russian // Proceedings of BSNLP workshop, ACL. — 2013. — P. 12–16.
 8. Chetviorkin I. I. , Loukachevitch N. V. Extraction of Russian Sentiment Lexicon for Product Meta-Domain // In Proceedings of COLING 2012: Technical Papers , pages 593–610]
 9. Chetviorkin I., Loukachevitch N. Sentiment analysis track at romip-2012 // компьютерная лингвистика и интеллектуальные технологии, по материалам конференции Диалог-2013,. — Vol. 2. — 2013. — P. 40–50.
 10. Das, S.R., Chen, M.Y.: Yahoo! For Amazon: Opinion Extraction From Small Talk On The Web. In: Proceedings Of The 8th Asia Pacific Finance Association Annual Conference. (2001)
 11. Dave, K.; Lawrence, S.; Pennock, D. M. 2003. Mining The Peanut Gallery: Opinion Extraction And Semantic Classification Of Product Reviews. In Proceedings Of Www-03, 12th International Conference On The World Wide Web, Acm Press, Budapest, Hu, 2003, Pp. 519–528
 12. Gamon, M. And A. Aue: 2005, ‘Automatic Identification Of Sentiment Vocabulary: Exploiting Low Association With Known Sentiment Terms’. // In: Proceedings Of The Acl-05 Workshop On Feature Engineering For Machine Learning In Natural Language Processing. Ann Arbor, Us.
 13. Gamon, M.: 2004, ‘Sentiment Classification On Customer Feedback Data: Noisy Data, Large Feature Vectors, And The Role Of Linguistic Analysis’. In: Proceeding Of Coling-04, The 20th International Conference On Computational Linguistics. Geneva, Ch, Pp. 841–847.
 14. Hatzivassiloglou, V., Wiebe, J.: Effects Of Adjective Orientation And Gradability On Sentence Subjectivity. In: 18th International Conference On Computational Linguistics (Coling-2000). (2000)
 15. Hu And Liu, 2004; Mining Opinion Features In Customer Reviews. In Proceedings Of Aaai-04, The 19th National Conference On Artificial Intelligence, San Jose, Us, 2004.

16. Hu, M., And Liu, B. 2004. Mining Opinion Features In Customer Reviews. To Appear In Aaai'04, 2004. In: Kdd-04. Pp. 168–177.
17. Hurst, M. And K. Nigam: 2004, 'Retrieving Topical Sentiments From Online Document Collection'. In: Exploring Attitude And Affect In Text: Theories And Application (Aaai-Eaat 2004). Stanford University.
18. Kamps And Marx, M.; Mokken, R.J.; De Rijke, M., 2004. Using Wordnet To Measure Semantic Orientation Of Adjectives. Lrec 2004, Volume Iv, Pp. 1115—1118
19. Kim And Hovy, 2004; 'Determining The Sentiment Of Opinions'. In: Proceedings Coling-04, The Conference On Computational Linguistics. Geneva, Ch, Pp. 1367–1373.
20. Kim, S.-M. And E. Hovy: 2005a, 'Automatic Detection Of Opinion Bearing Words And Sentences'. // In: Companion Volume To The Proceedings Of Ijcnlp-05, The Second International Joint Conference On Natural Language Processing. Jeju Island, Kr, Pp. 61–66.
21. Kim, S.-M. And E. Hovy: 2005b, 'Identifying Opinion Holders For Question Answering In Opinion Texts'. In: Proceedings Of Aaai-05 Workshop On Question Answering In Restricted Domains. Pittsburgh, Us.
22. Kim, S.-M. And E. Hovy: 2006, 'Extracting Opinions, Opinion Holders, And Topics Expressed In Online News Media Text'. In: Proceedings Of The Acl/Coling Workshop On Sentiment And Subjectivity In Text. Sydney, Australia.
23. Kobayashi Et Al., 2007
24. *Kuznetsova E., Loukachevitch N., Chetviorkin I.* Testing rules for a sentiment analysis system // *омпьютерная лингвистика и интеллектуальные технологии, по материалам конференции Диалог-2013.*, — Vol. 2. — 2013. — P. 71–80.
25. Mishne G. And M. De Rijke, 2006. "Capturing Global Mood Levels Using Blog Posts," In: Computational Approaches To Analyzing Weblogs: Papers From The 2006 Aaai Spring Symposium. Menlo Park, Calif.: Aaai Press, Pp. 145–152, And At [Http://Staff.Science.Uva.Nl/~Gilad/Pubs/Aaai06-Blogmoods.Pdf](http://Staff.Science.Uva.Nl/~Gilad/Pubs/Aaai06-Blogmoods.Pdf), Accessed 21 August 2007.
26. Mishne G. And N. Glance, 2006. "Leave A Reply: An Analysis Of Weblog Comments," Proceedings Of The 3rd Annual Workshop On The Weblogging Ecosystem: Aggregation, Analysis And Dynamics, 15th World Wid Web Conference, At [Http://Www.Blogpulse.Com/Www2006-Workshop/Papers/Wwe2006-Blogcomments.Pdf](http://Www.Blogpulse.Com/Www2006-Workshop/Papers/Wwe2006-Blogcomments.Pdf), Accessed 21 August 2007.

27. Mulder, M.; Nijholt, A.; Den Uyl, M.; Terpstra, P., A Lexical Grammatica Implementation Of Affect, Proceedings Of Tsd-04, The 7th International Conference Text, Speech And Dialogue, Lecture Notes In Computer Science, Vol. 3206, Springer-Verlag, Brno, Cz, 2004, Pp. 171–178.
28. Pang, B., Lee, L., Opinion Mining And Sentiment Analysis, 2008
29. Pang, B., Lee, L., Vaithyanathan, S.: Thumbs Up? Sentiment Classification Using Machine Learning Techniques. In: Proceedings Of The Conference On Empirical Methods In Natural Language Processing (Emnlp-2002). (2002)
30. Popescu, A; Etzioni, O; 2005. Extracting Product Features And Opinions From Reviews. In Proceedings Of Hlt-Emnlp-05, The Human Language Technology Conference/Conference On Empirical Methods In Natural Language Processing, Vancouver, Ca, 2005.
31. Riloff, E., S. Patwardhan, And J. Wiebe: 2006, 'Feature Subsumption For Opinion Analysis'. In: Proceedings Of Emnlp-06, The Conference On Empirical Methods In Natural Language Processing. Sydney, Aus, Pp. 440–448.
32. Riloff, E., Wiebe, J.: Learning Extraction Patterns For Subjective Expressions. In: Proceedings Of The Conference On Empirical Methods In Natural Language Processing (Emnlp-2003). (2003) 105–112
33. Riloff, E.; Wieber, J.; Wilson, T. Learning Subjective Nouns Using Extraction Pattern Bootstrapping. In Walter Daelemans And Miles Osborne, Editors, Proceedings Of Conll-03, 7th Conference On Natural Language Learning, Edmonton, Ca, 2003, Pp. 25–32.
34. Taboada, M.; Grieve, J. Analyzing Appraisal Automatically. In Proceedings Of The Aaai Spring Symposium On Exploring Attitude And Affect In Text: Theories And Applications, Stanford, Us, 2004
35. Tong, "An Operational System For Detecting And Tracking Opinions In On-Line Discussion," In Proceedings Of The Workshop On Operational Text Classification (Otc), 2001.
36. Turney, P., Littman, M.: Measuring Praise And Criticism: Inference Of Semantic Orientation From Association. *Acm Transactions On Information Systems (Tois)* 21 (2003) 315–346
37. Turney, P., Thumbs Up Or Thumbs Down? Semantic Orientation Applied To Unsupervised Classification Of Reviews, Proceedings Of Acl-02, 40th Annual Meeting Of The Association For Computational Linguistics, Association For Computational Linguistics, 2002, Pp. 417–424.)

38. Wiebe, J. And E. Riloff: 2005, 'Creating Subjective And Objective Sentence Classifiers From Unannotated Texts'. In: Proceeding Of Cicling-05, International Conference On Intelligent Text Processing And Computational Linguistics, Vol. 3406 Of Lecture Notes In Computer Science. Mexico City, Mx, Pp. 475–486.
39. Wiebe, J., E. Breck, C. Buckley, C. Cardie, P. Davis, B. Fraser, D. Litman, D. Pierce, E. Riloff, T. Wilson, D. Day, And M. Maybury: 2003, 'Recognizing And Organizing Opinions Expressed In World Press'. In: Proceedings Of The Aa Ai Spring Symposium On New Directions In Question Answering.
40. Wiebe, J.: 2002, 'Instructions For Annotating Opinions In Newspaper Artic'. Technical Report Tr-01-101, University Of Pittsburgh, Department Of Computer Science, Pittsburgh, Pa.
41. Wiebe, J.: Learning Subjective Adjectives From Corpora. In: Proceedings Of The Seventeenth National Conference On Artificial Intelligence (Aaai-2000). (2000) 735–740
42. Wilson Theresa, Janycewiebe, Paul Hoffmann. Recognizing Contextual Polarity In Phrase-Level Sentiment Analysis.
43. Yi, J., Nasukawa, T., Bunescu, R., Niblack, W.: Sentiment Analyzer: Extracting Sentiments About A Given Topic Using Natural Language Processing Techniques. In: Proceedings Of The 3rd Ieee International Conference On Data Mining (Icdm-2003). (2003)
44. Антонова, Соловьев. Использование метода условных случайных полей для обработки текстов на русском языке // <http://www.dialog-21.ru/digests/dialog2013/materials/pdf/AntonovaAJ.pdf>
45. Апресян Ю.Д. Лексическая семантика: синонимические средства языка. М.: Наука, 1974.
46. Арутюнова Н.Д. Типы языковых значений: Оценка. Событие. Факт – М.: Наука, 1988
47. Будаев Э. В., Чудинов А. П. Современная политическая лингвистика [Электронный ресурс]. URL: <http://philology.ru/linguistics1/budaev-chudinov-06a.htm> (дата обращения: 12.10.2010).
48. Васильев В. Г., Худякова М. В., Давыдов С. Классификация отзывов пользователей с использованием фрагментарных правил // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая - 3 июня 2012г.).

- Вып. 11 (18).- М.: Изд-во РГГУ, 2012.
<http://www.dialog-21.ru/digests/dialog2012/materials/pdf/132.pdf>
49. Вольф Е.М. Функциональная семантика оценки. М.: Наука, 1985
50. Демьянков, В.З. Политический дискурс как предмет политологической филологии / В.З Демьянков // Политическая наука. – 2002. - №3.- С.32
51. Ермаков А.Е., Киселев С.Л. Лингвистическая модель для компьютерного анализа тональности публикаций СМИ. // Компьютерная лингвистика и интеллектуальные технологии: труды международной конференции Диалог'2005. – Москва, Наука, 2005.
52. Карпова О.С., Резникова Т.И., Рахилина Е.В., Рыжова Д.А. Оценочные значения ребрендингового типа в признаковой лексике (по материалам Базы данных семантических переходов в качественных прилагательных и наречиях) // Компьютерная лингвистика и интеллектуальные технологии. Вып. 10 (17). М.: Изд-во РГГУ, 2011, 292-304.
53. Кашкин Е., Ляшевская О. Н. Семантические роли и сеть конструкций в системе FrameBank // В кн.: Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 29 мая - 2 июня 2013 г.). В 2-х т. Т. 1: Основная программа конференции. Вып. 12 (19). М. : РГГУ, 2013. С. 325-343.
54. Кобозева И.М. Лингвистическая семантика: Учебник. Изд. 4-е. М.: Книжный дом «ЛИБРОКОМ», 2009
55. Котельников Е. В., Клековкина М. В. Автоматический анализ тональности текстов на основе методов машинного обучения // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая - 3 июня 2012г.). Вып. 11 (18).- М.: Изд-во РГГУ, 2012.
<http://www.dialog-21.ru/digests/dialog2012/materials/pdf/105.pdf>
56. Кравченко А. Н. Генерация шаблонов оценочных выражений на основе размеченного текста // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая - 3 июня 2012г.). Вып. 11 (18).- М.: Изд-во РГГУ, 2012.
<http://www.dialog-21.ru/digests/dialog2012/materials/pdf/129.pdf>

57. Кустова Г.И. Оценка как универсальная категория и типы оценочных значений прилагательных // Материалы конференции "Проблемы лексико-семантической типологии (ВГУ, 8-9 октября 2010 г.)", Воронеж, 2011. С. 246-253.
58. Кустова Г.И. Производные значения с экспериенциальной составляющей // Семиотика и информатика, вып. 36. 1998.
59. Кустова Г.И. Словарь русской идиоматики. Сочетания слов со значением высокой степени. М., 2008 (<http://dict.ruslang.ru/magn.php>).
60. Кустова Г.И. Типы производных значений и механизмы языкового расширения. М.: Языки славянской культуры, 2004.
61. Лукашевич Н., Четверкин И. Извлечение и использование оценочных слов в задаче классификации отзывов на три класса // Вычислительные методы и программирование: Новые вычислительные технологии (Электронный научный журнал). — 2011. — С. 73–81. В статье рассматривается задача извлечения оценочных слов из отзывов пользователей на основе большого количества признаков. Далее исследуется вопрос об использовании этих оценочных слов в задаче классификации отзывов на три класса.
62. Лукашевич Н., Четвёркин И. Построение модели для извлечения оценочной лексики в различных предметных областях // Моделирование и анализ информационных систем. — 2013. — Т. 20, № 2. — С. 70–79.
63. Мартыненко Г. Я., Чебанов С. В. Стилеметрия // Прикладная лингвистика. СПб, 1996.
64. Маслова В. А. Политическая лингвистика. Екатеринбург, 2008. С. 43-48.
65. Пазельская, Соловьев. Метод определения эмоций в текстах на русском языке // <http://www.dialog-21.ru/digests/dialog2011/materials/en/pdf/50.pdf>.
66. Поляков П. Ю., Калинина М. В., Плешко В. В. Исследование применимости методов тематической классификации в задаче классификации отзывов о книгах // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая - 3 июня 2012г.). Вып. 11 (18).- М.: Изд-во РГГУ, 2012. <http://www.dialog-21.ru/digests/dialog2012/materials/pdf/103.pdf>
67. Толдова С. Ю., Кустова Г. И., Ляшевская О. Н. Семантические фильтры для разрешения многозначности в национальном корпусе русского языка: глаголы // В кн.: Компьютерная лингвистика и интеллектуальные технологии: По

материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2008 г.) Вып. 7. М. : РГГУ, 2008. С. 522-529.

68. Толдова С. Ю., Ляшевская О. Н. Современные проблемы и тенденции компьютерной лингвистики // Вопросы языкознания. 2014. № 1. С. 120-145.

Список источников:

1. Абрамов Н. Словарь русских синонимов и сходных по смыслу выражений: Около 5 000 синонимических рядов. Более 20 000 синонимов — 7-е изд., стереотип. — М.: Русские словари, 1999.
2. Национальный корпус русского языка: www.ruscorpora.ru.
3. Интернет-ресурс www.livejournal.com.
4. Интернет-ресурс www.echo.msk.ru.
5. Система фреймов семантических ролей Frambank: www.framebank.ru.
6. Словарь русской идиоматики Г.И. Кустовой: www.dict.ruslang.ru/magn.php.
7. Сайт Российского семинара по оценке методов информационного поиска: www.romip.ru

Приложение 1 (электронное). Список переносящих оценку фреймов с ролью результата.

Приложение 2 (электронное). Словарь оценочной лексики.

Приложение 3 (электронное). Корпус политических блогов, размеченный по явной и скрытой оценке.