

## Определение тональности сообщений СМИ методом их концептуального анализа

**Хорошилов Ал-др А. \***

ФИЦ ИУ РАН, Москва, Россия  
khorosilov@mail.ru

**Козловская Я.Д. \*\***

МАИ, Москва, Россия  
yana04029877@mail.ru

**Мусабаев Р.Р. \*\*\***

ИИВТ, Алматы, Казахстан  
rmusab@gmail.com

**Красовицкий А.М. \*\*\*\***

ИИВТ, Алматы, Казахстан  
akrassovitskiy@gmail.com

**Хорошилов Ал-ей А. \*\*\*\*\***

«27 ЦНИИ» Минобороны России, Москва, Россия  
alex\_khor@mail.ru

В статье изложено решение задачи создания лингвистического инструментария и методики автоматического определения тональности новостных сообщений, связанных с качеством жизни рядового гражданина республики Казахстан.

### Для цитаты:

*Хорошилов Ал-др А., Козловская Я.Д., Мусабаев Р.Р., Красовицкий А.М., Хорошилов Ал-ей. А. Определение тональности сообщений СМИ методом их концептуального анализа. // Моделирование и анализ данных. 2019. Том 09. № 4. С. 67–79. doi: 10.17759/mda.2019090405*

\***Хорошилов Александр Алексеевич**, доктор технических наук, ведущий научный сотрудник Федерального исследовательского центра «Информатика и управление» Российской академии наук (ФИЦ ИУ РАН), профессор кафедры Московского авиационного института (МАИ), Москва. E-mail: khorosilov@mail.ru

\*\***Козловская Яна Дмитриевна**, студент Московского авиационного института (национальный исследовательский университет), Россия. E-mail: yana04029877@mail.ru

\*\*\***Мусабаев Рустам Рафикович**, руководитель лаборатории Института информационных и вычислительных технологий, Казахстан. E-mail: rmusab@gmail.com

\*\*\*\***Красовицкий Александр Михайлович**, кандидат технических наук, ведущий научный сотрудник Института информационных и вычислительных технологий, Казахстан. E-mail: akrassovitskiy@gmail.com

\*\*\*\*\***Хорошилов Алексей Алексеевич**, к.т.н., старший научный сотрудник 27-го Центрального научно-исследовательского института Министерства обороны России. E-mail: alex\_khor@mail.ru



Был определен подход для решения задачи, разработано программное обеспечение и методы автоматизированного создания словарей объектов и словарей оценочных предикатов, а также словарей модификаторов меры оценки. Проведенный эксперимент подтвердил правильность предложенной методики оценки событий, освещенных в новостных сообщениях и работоспособность программного комплекса.

Эта методика, при соответствующей выборке объектов оценки событий, может быть использована при создании тональных портретов конкретных авторов по совокупности их публикаций, а также тональных портретов различных новостных агрегатов по совокупности освещаемых ими событий в конкретный временной интервал.

**Ключевые слова:** концептуальный анализ, предикатно-актантная структура, тональность текстов, тональный словарь, модификаторы тональности, автоматическое установление оценки тональности текстов.

## Введение

На психоэмоциональной настрой современного общества значительное влияние оказывают электронные средства массовой информации (ЭСМИ). Эти средства информируют общество о событиях, происходящих в мире. Но, как правило, большинство событий освещается с разных точек зрения и в значительной части случаев цель такого освещения событий является желание автора публикации сформировать у читателя определенной мнение, созвучное мнению автора. Но самом деле ЭСМИ в ряде случаев навязывают такое мнение читателям, которое заказывается авторам идеологических публикаций. Как правило, чтобы сформировать у читателя определенное мнение по идеологическим вопросам читателю преподносится заранее препарированная информация – подаются определенные события в позитивном или негативном плане, или некоторые события или вообще не происходили (так называемые фейковые события), или являлись постановочными событиями. Такое освещение событий формирует определенный психоэмоциональный настрой у определенной части общества. Сейчас все такие манипуляции с ЭСМИ называют элементами информационной войны. Для борьбы с такими манипуляциями необходимо выработать автоматизированные средства противодействия, базирующиеся на методах искусственного интеллекта.

Одним из наиболее действенных способов противодействия этим манипуляциям был разработан в США в начале Второй мировой войны, когда Германия объявила войну США. В то время в США были сильны профашистские настроения и функционировало большое число газет, пропагандирующих эти настроения. Поскольку США являлось демократическим правовым государством, то необходимы были формальные основания для закрытия по судебному решению этих профашистских газет. Таким основанием могли бы быть количественные данные по фактам позитивного отношения к фашистской идеологии и одновременно негативного – к устоям демократического общества. Для решения этой задачи были привлечены лингвисты, которые проанализировали большое число газетных публикаций и установили критерии, в соответствии можно было бы формально отнести газеты к профашистским. При этом была предложена относительная простая методика, заключающаяся в выборе списка значимых фамилий и терминов, относящихся как к фашистской идеологии, так и к демократическим ценностям, с целью определения тональности контента всех случаев упоминания этих терминов и фамилий. Впоследствии такой метод определения тональности текстов назвали контент-анализом.



## **Обзор методов определения тональности текстов**

При решении проблемы автоматического установления тональности текстов можно выделить несколько основных подходов, изложенных в работах [1–5]:

Подходы, базирующиеся на методах машинного обучения и векторном сравнении исходных текстов с размеченными эталонными коллекциями, предполагают предварительное обучение машинного классификатора путем раскладывания каждого документа в виде вектора признаков. При этом каждому документу указывается его тип тональности. При реализации этих методов требуются значительные первоначальной трудозатраты при разметке текстов, необходимые для обучения алгоритмов обучения. Эти методы достаточно эффективны при анализе однородных текстов слабо подверженным изменениям тематики и их эмотивного настроения.

Подходы, базирующиеся на выявлении в текстах эмотивной лексики на основе использования тональных словарей. В рамках этого метода часто используются списки паттернов, подставляемых в регулярные выражения, а также правила формирования шаблонов для объединения слов в цепочки. Обычно в качестве шаблонов использовались два типа синтаксических конструкций: «прилагательное + существительное» и «наречие + глагол». При этом тональность текста рассчитывалась как среднее эмоциональных оценок. Оценка для каждой конструкции определяется путем установления совместной встречаемости их со словами тонального словаря.

С нашей точки зрения наиболее перспективными и универсальными подходами к решению задачи установления тональности текстов являются подходы, ориентированные на глубокий семантический анализ текстов. В рамках этого подхода возможно обеспечить детальный анализ тональности в пределах каждого предложения и в отношении каждого объекта. Так, например, в работе [1] производилась определения тональности у объектов на основе предикационных отношений в пропозиции. Тональность устанавливалась на основе классификационных тональных словарей.

В работе [5] ставилась задача разработки методов оценки деятельности органов власти на основе набора показателей видов их деятельности и процедурных средств лингвистического анализа текстов. Преимущество таких методов заключается в том, что они обеспечивают получение более объективных количественных показателей и освобождают результаты оценки от субъективного экспертного представления о действительной ситуации.

## **Постановка задачи исследований**

Исходя из конкретных задач НИР<sup>1</sup> в рамках нашего исследования была поставлена задача разработки лингвистического инструментария и методики автоматического определения тональности новостных сообщений, связанных с качеством жизни<sup>2</sup> рядового гражданина. Как известно, информация, относящаяся к повышению или понижению качества жизни, в значительной степени определяют его эмоциональное отношение к властным структурам. При этом мы исходили из того, что на эмоциональный настрой пользовате-

<sup>1</sup> Работа проводится в рамках ПЦФ BR05236839 «Разработка информационных технологий и систем для стимулирования устойчивого развития личности как одна из основ развития цифрового Казахстана».

<sup>2</sup> Совокупность свойств и характеристик жизни человека, относящихся к их способности удовлетворять его существующие и предполагаемые потребности [ГОСТ Р 22.10.01–2001].



лей ЭСМИ влияет сознательный подбор освещаемых событий и их эмоциональная окраска. Для решения этой задачи необходимо исследовать содержательную и эмоциональную составляющие текстов, составить оценочные словари и разработать программные средства их формализации и выявления в текстах.

Для установления количественных характеристик тональности текстов воспользуемся формулой подсчета значения веса тональности текстов, приведенной в работе [2], но для нормирования числа предложений в документе берем не их абсолютное значение, а логарифм значения:

$$W_k = \frac{\sum_{i=1}^{N_c} W_i}{\log(N)}, \quad (1)$$

где

$W_k$  – значение веса тональности  $k$ -ого документа;

$W_i$  – значение веса тональности  $i$ -ой ситуации;

$N$  – общее число предложений в документе.

Для формирования тонального портрета подмножества документов все тексты  $T_i$  помещаются в одномерное эмотивное пространство в соответствии со своим весом  $W_{T_i}$ , причем тексты положительной тональности должны занимать в этом пространстве местоположение справа, отрицательной тональности – местоположение слева. Тексты нейтральной тональности должны располагаться в центре этого пространства. Среднее значение текстов положительной или отрицательной тональности может быть вычислено по формуле:

$$AW_t^c = \frac{\sum_{i=1}^{N_c} W_{T_i}}{N_c}, \quad T_i \in C, \quad (2)$$

где,

$AW_t^c$  – средний вес текстов, относящихся к классу тональности  $C$ ;

$N_c$  – число текстов относящихся к классу тональности  $C$ ;

После вычисления средних значений текстов положительной или негативной тональности необходимо установить для них границу  $D$  – как середину отрезка  $[AW_t^- ; AW_t^+]$ . На этом отрезке текст  $T$  в соответствии со его знаком и значение его веса располагаются справа или слева от границы  $D$ . При этом, как уже было отмечено, все тексты справа от границы  $D$  относятся к положительному классу, а слева – к отрицательному.

## Теоретические предпосылки решения задачи

Общеизвестно, что смысловое содержание текстов или их фрагментов выражается через систему единиц смысла и их отношений. Наиболее устойчивыми единицами смысла являются понятия (отдельные слова и устойчивые фразеологические и терминологические словосочетания). Второй по значимости единицей смысла является предложение. Основной чертой предложений является их предикативность – то есть свойство утверждения наличия у объектов определенных признаков и их отношений. В основе смысловой структуры предложений лежит его предикатно-актантная структура, компонентами которой являются понятия-предикаты (признаки и отношения) и понятия-актанты, выступающие в роли описываемых объектов.

Для установления психоэмоционального настроения текстов необходимо определить объекты, в отношении которых производится оценка и те компоненты смысловой структуры, которые определяют такую оценку. В качестве объектов обычно выступают суще-



ствительные, в качестве определителей оценки – глаголы. Для усиления или уменьшения эмоциональной оценки используются прилагательные и наречия. Для более точной оценки тональности текстов были предложены три категории объектов. В качестве таких объектов первой категории фигурируют, все те объекты, явления и ситуации, которые в наибольшей степени влияют на качество жизни гражданина. В качестве предикатов – те процессы, которые улучшают или ухудшают это качество. Примерами таких объектов и их оценок могут служить доступность и стоимость предметов первой необходимости (продуктов питания, медикаментов, одежды и др.), качество здравоохранения, социальной защиты, ЖКХ и борьба с правонарушениями. В качестве объектов второй категории – все те ситуации, которые очень важны для общества в целом, но в меньшей степени влияют на качество жизни конкретного гражданина. К таким явлениям можно отнести борьбу с коррупцией, борьбу с терроризмом, правоохранительную, судебную и налоговую систему. И, наконец, к третьей категории можно отнести все те явления, которые определяют престиж страны в целом. Это международная деятельность, культура, спорт, торговля, отношения с другими странами и др.

### **Методика и технологии создание декларативных средств**

При разработке методов оценки текстов необходимо разработать эталонные оценочные словари и методику их применения при анализе текстов. В рамках наших исследований мы пытались решить эту проблему путем автоматизации процесса формирования ситуативных оценочных словарей. Как следует из работ [2] для определения ситуаций и их тональной оценки требуется создать комплекс словарей, состоящий из объектов и связанных с ними предикатов, а также их определителей – наречий (для глаголов) и прилагательных (для существительных), выступающих в роли модификаторов их меры тональности. Обычно решение этой задачи возлагалось на экспертное сообщество. При таком решении затрачивались значительные человеческие ресурсы и длительное время на их создание. Автоматизация этого процесса позволит сократить трудозатраты и время на создание таких словарей.

Основная идея метода автоматизированного составления тональных словарей заключается в том, чтобы на основе уже имеющихся словарных ресурсов обеспечить возможность автоматического выявления в текстах объектов оценки. К таким словарным ресурсами можно отнести различные перечни объектов и текстов, в которых они встречаются. Процесс выявления объектов первой необходимости можно начать, например, с ассортимента потребительской корзины<sup>3</sup>. Этот ассортимент товаров и услуг необходимо выявить в текстах путем их лингвистического анализа и установить в текстах их смысловую связь с лексическими конструкциями, осуществляющими оценку объектов. В предложении такими конструкциями являются глаголы (предикаты). Эксперту должен для каждой такой пары «объект-предикат» зафиксировать оценочный глагол и назначить этой паре индекс тональности (позитив/негатив). Категория тональности для этих объектов будет устанавливаться автоматически. Далее, на основе установленных предикатов, можно будет опре-

<sup>3</sup> «Необходимые для сохранения здоровья человека и обеспечения его жизнедеятельности минимальный набор продуктов питания, а также непродовольственные товары и услуги, стоимость которых определяется в соотношении со стоимостью минимального набора продуктов питания» [Потребительская корзина 2013–2018 год. Информационный портал РФ].



делить в коллекции текстов все возможные объекты, встречающиеся совместно с ними. Эксперту должен только установить индекс их тональности. Объекты других категорий, относящиеся к отраслевым видам деятельности, можно будет установить по различным отраслевым документам: глоссариям, тезаурусам, онтологиям, нормативно-справочным документам и др. Здесь также задача эксперта для каждой выявленной пары «объект-предикат» установить индекс тональности. Категория будет устанавливаться автоматически в соответствии с установленной категории для данной отрасли.

Модификаторы степени тональности: наречия (для глаголов) и прилагательные (для существительных) устанавливаются на основе результатов семантико-синтаксического анализа. Задача экспертов также сводится к установлению степени тональности (слабая/средняя/сильная).

После того как выполнен первоначальный этап установления объектов, предикатов и модификаторов, относящихся к оценке тональности объектов первой категории, можно будет перейти к реализации второго этапа, целью которого будет выявление объектов второй и третьей категории и их оценочных конструкций. Этот этап по своим задачам гораздо сложнее, поскольку необходимо определить виды деятельности в отраслях и установить наборы их оценочных показателей. Поэтому этот этап разработки должен базироваться в большей степени на технологиях глубинного семантического анализа текстов, позволяющего автоматически выявить в отраслевых текстах компоненты их предикатно-актантной структуры и автоматически назначить им набор тональных признаков. В табл.1 приводятся фрагменты комплекса оценочных тональных словарей.

Таблица 1

Категория	Компоненты предикатно-актантной структуры	Классификационный индекс	Прямая/Обратная совместимость	Количественная характеристика
<b>Основные объекты (виды деятельности в отраслях)</b>				
1	<i>хлебобулочные изделия</i>	<i>N00001.001</i>	<i>D</i>	<i>1.0</i>
1	<i>вино-водочная продукция</i>	<i>N00001.001</i>	<i>D</i>	<i>1.0</i>
1	<i>куриные яйца</i>	<i>N00001.001</i>	<i>D</i>	<i>1.0</i>
2	<i>налог на имущество</i>	<i>N00254.001</i>	<i>D</i>	<i>0.6</i>
2	<i>подходящий налог</i>	<i>N00254.001</i>	<i>D</i>	<i>0.6</i>
3	<i>футбольная команда</i>	<i>N03675.001</i>	<i>D</i>	<i>0.3</i>
<b>Оценочные предикаты (предикативные конструкции)</b>				
0	<i>подешеветь</i>	<i>V00042.001</i>	<i>D</i>	<i>1.0</i>
0	<i>подорожать</i>	<i>V000044.001</i>	<i>N</i>	<i>1.0</i>
0	<i>увеличился</i>	<i>V000048.001</i>	<i>N</i>	<i>1.0</i>
0	<i>понижился</i>	<i>V000049.001</i>	<i>D</i>	<i>1.0</i>
0	<i>выиграть</i>	<i>V000021.001</i>	<i>D</i>	<i>1.0</i>
0	<i>проиграть</i>	<i>V000520.001</i>	<i>D</i>	<i>1.0</i>
<b>Модификаторы тональности – прилагательные</b>				
0	<i>незначительный</i>	<i>A000052.001</i>	<i>D</i>	<i>0.8</i>
0	<i>весомый</i>	<i>A000021.001</i>	<i>D</i>	<i>1</i>
<b>Модификаторы тональности – наречия</b>				
0	<i>очень</i>	<i>Y000042.001</i>	<i>D</i>	<i>1.2</i>
0	<i>незначительно</i>	<i>Y000037.001</i>	<i>D</i>	<i>0.8</i>



В табл. 3 для каждого словарного объекта приводится индекс его категории, классификационный индекс, индекс прямой или обратной совместимости и весовой коэффициент. Классификационный индекс объекта соответствует его родовому индексу в тематической онтологии, а индекс прямой совместимости сочетания «объекта- предикат» соответствует основной тональности предиката, или обратный индекс – изменяет эту тональность на противоположную.

В табл. 2 приведены количественные характеристики словарей текущей версии словарного комплекса.

Таблица 2

n/n	Наименование словаря	Объем словаря
Словарь объектов (видов деятельности в отраслях)		
1	Словарь «Продуктовая корзина»	897
2	Словарь по ЖКХ	3427
3	Словарь по спорту	2588
4	Налоговый словарь	2764
5	Словарь по международной деятельности	3536
6	Словарь оценочных предикатов	1736
7	Словари модификаторов тональности: прилагательных	237
8	наречий	147
	<b>Общее количество словарных статей</b>	<b>15332</b>

## Технология автоматического установления тональности текстов

Технология автоматического установления тональности текстов базируется на процедурах их семантико-синтаксического анализа, в результате работы которой смысловое содержание текстов формализуется и структурируется, и выявляются компоненты предикатно-актантной структуры. Эти формализованные компоненты последовательно сопоставляются с комплексом тональных словарей: вначале сопоставляются наименования объектов, после этого, в случае их успешного поиска в словарях объектов, производится сопоставление выявленных в тексте предикатов с элементами словаря оценочных предикатов. При успешном сопоставлении и при наличии в текстах предполагаемых модификаторов меры тональности (наречий и прилагательных) также производится их поиск в соответствующих словарях. Далее после получения всей необходимой информации о событии вычисляется его количественная характеристика по следующей формуле:

Технологию автоматического установления тональности текстов можно представить в виде следующей последовательности лингвистических процедур обработки текста:

1. Выполнение формально-логического контроля текста.
2. Членение на отдельные слова (по пробелам и разделительным знакам между ними).
3. Выполнение морфологического анализа слов.
4. Выполнение членения текста на предложения.
5. Выполнение семантико-синтаксического анализа предложений.
6. Выполнение концептуального анализа предложений и выявление в них компонентов предикатно-актантной структуры.



7. Приведение компонентов предикатно-актантной структуры к их формализованному представлению.
8. Сопоставление компонент предикатно-актантной структуры с элементами комплекса тональных оценочных словарей.
9. Вычисление количественных характеристик тональности событий.
10. Формирование статистического отчета о тональности текста.

## Программные средства определения тональности текстов

При реализации программно-технологического комплекса определения тональности текстов мы базировались на процедурах программно-лингвистической платформе MetaFraz [6], в рамках которой выполнялась формализация смыслового содержания текстов и выявлялась его формализованная предикатно-актантная структура. Методы и технологии машинной грамматики и семантико-синтаксического анализа достаточно подробно описаны в работах [7,8]. Разработанный в рамках данного исследования программный модуль осуществлял разбор на отдельные компоненты предикатно-актантную структуру предложений и производил сопоставление по определенной схеме этих компонент с тональными оценочными словарями. Аналогично сопоставлялись с соответствующими словарями модификаторами меры тональности – наречия и прилагательные. Программно-технологическая схема подсистемы автоматического определения тональности текстов приведена на рис.1. (см. стр. 9)

## Методика оценки тональности текстов

Методика оценки тональности текстов должна в какой-то степени моделировать восприятие человеком смыслового содержания новостных сообщения и их эмоциональной окраски. Обычно он воспринимает всю сумму событий, содержащихся в этих сообщениях, и формирует свое собственное мнение об этих событиях. Это мнение во многом зависит, от того насколько эти события влияют или могут повлиять на его качество жизни и в гораздо меньшей степени на его мнения может повлиять мнение автора публикации. Поэтому мы согласились с авторами статьи [2] утверждавших, что лучшие результаты были получены на основе суммы эмоциональных оценок отдельных сообщений, а не на их частоту употребления, как обычно это применяется в методах машинного обучения.

Количественное значение тональной оценки  $i$ -ого события вычислялось по формуле:

$$S_i = C_i \cdot K_{oi} \cdot K_{mi} \cdot K_{ni},$$

где

$S_i$  — значение тональной оценки  $i$ -ого события;

$C_i$  —  $i$ -ое событие;

$K_{oi}$  — коэффициент категории  $i$ -ого объекта события (диапазон значений коэффициента [-1, -0.6, -0.3, +0.3, +0.6, ±1]);

$K_{mi}$  — коэффициент  $i$ -ого модификатора оценочного предиката (диапазон значений коэффициента [0.1, 1.2]);

$K_{ni}$  — коэффициент  $i$ -ого модификатора объекта события (диапазон значений коэффициента [0.1, 1.2]).

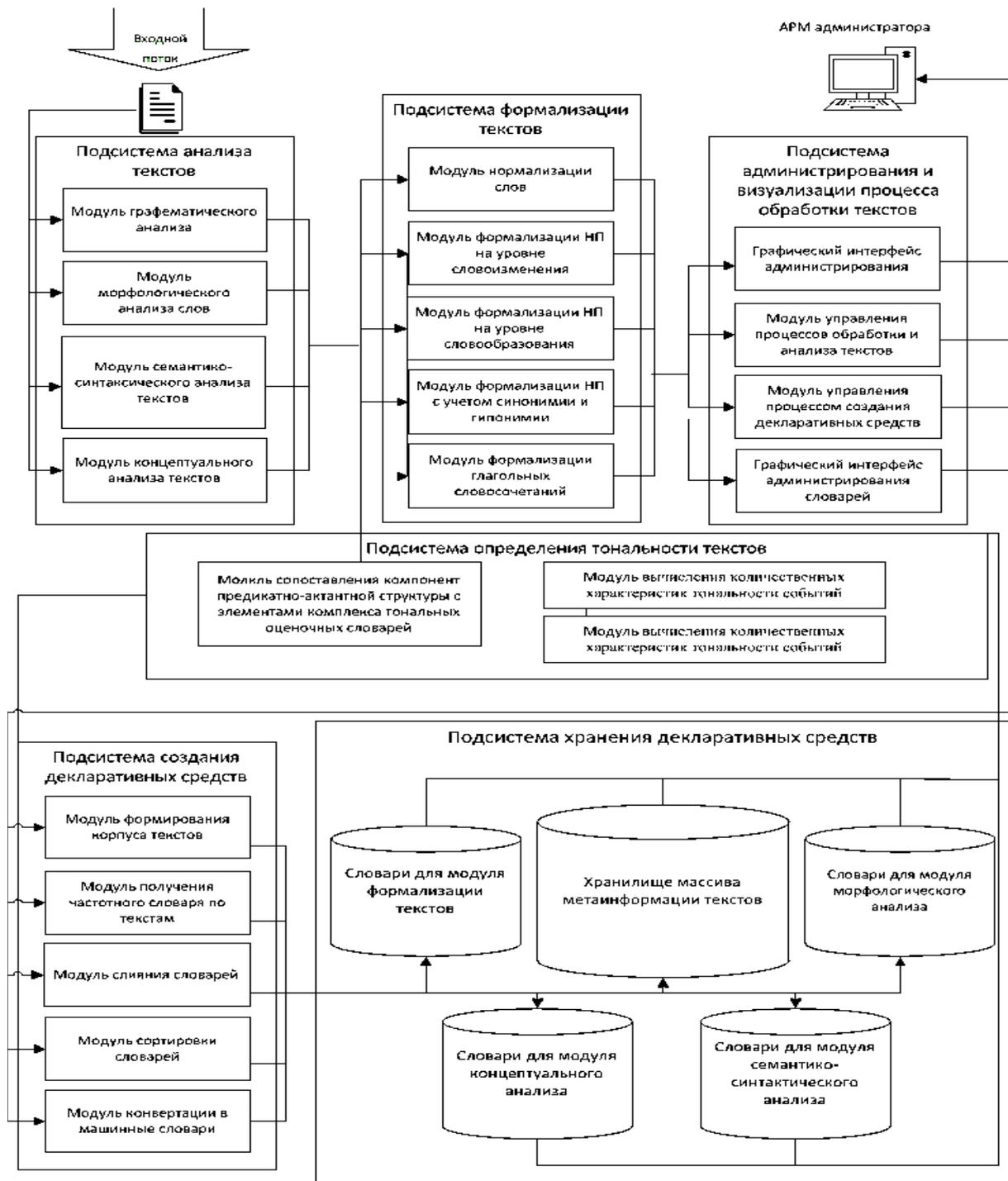


Рис.1. Программно-технологическая схема подсистемы автоматического определения тональности текстов.

## Эксперимент по автоматическому установлению тональности текстов

Задача эксперимента заключалась в проверке работоспособности предложенных технологических решений и проведения экспертной оценки результатов автоматического установления оценки тональности текстов. Для каждого документа выявлялись все ана-



лизируемые ситуации и вычислялась для каждой ситуации его тональная оценка, далее вычислялась абсолютная и средняя тональная оценка документа ( $W$ ). В табл. 3 приведены результаты автоматической оценки тональности документов, влияющих на качество жизни рядового гражданина.

Таблица 3

	Название документа и значение его тональной оценки (Т)	Позитивная оценка			Негативная оценка		
		3-я категория	2-ая категория	1-ая категория	1-ая категория	2-ая категория	3-я категория
1	Какие продукты подорожали в Казахстане ( $W=-1.49$ )			12	14		
2	Хорошее образование не может быть дешевым ( $W=1.34$ )		4			1	
3	Для большинства казахстанцев девальвация – это благо ( $W=0.91$ )		2	1			
4	Великая Победа – напоминание всем тем, кто вынашивает планы против нас ( $W=0.60$ )		3			1	
5	Проблемы образования касаются каждого ( $W=0.83$ )		3				
6	Хорошее образование не может быть дешевым ( $W=1.49$ )		4				
7	Чем обернется «дедолларизация» в СНГ, предложенная Владимиром Путиным ( $W=0.32$ )		3			2	
8	Празднование 550-летия Казанского ханства должно поднять наш регион на новый уровень ( $W=1.52$ )	6	1				

### Перспективы развития метода концептуального анализа текстов

Перспективы развития метода концептуального анализа текстов, ориентированного на автоматическое установление тональности событий, освещаемых в текстах, может развиваться по нескольким направлениям:

1. Совершенствованиям алгоритмов семантико-синтаксического и концептуального анализа текстов связано с повышением точности лексического анализа, разрешением неоднозначной синтаксических и семантических ситуаций и широким применением словарей синтаксических конструкций и системы правил их трансформаций [8].
2. Разработка методов выявления объектов анализа и их оценки должна базироваться на масштабных моделях предметных областей, представленных в виде низшего и среднего уровня тематических онтологии и включающих основной понятийный состав предметных областей и систему смысловых отношений между наименованиями понятий этих предметных областей [5].
3. Совершенствование методов оценки текстовых событий связаны с моделированием восприятия человеком позитивных или негативных событий и формирования у него соответствующего психоэмоционального настроения к причинам, вызывающим эти события [5].



## Заключение

В статье изложено решение задачи создания лингвистического инструментария и методики автоматического определения тональности новостных сообщений, связанных с качеством жизни рядового гражданина. Был определен подход для решения задачи, разработано программное обеспечение и методы автоматизированного создания словарей объектов и словарей оценочных предикатов, а также словарей модификаторов меры оценки. Проведенный эксперимент подтвердил правильность предложенной методики оценки событий, освещенных в новостных сообщениях и работоспособность программного комплекса.

Эта методика, при соответствующей выборке объектов оценки событий, может быть использована при создании тональных портретов конкретных авторов по совокупности их публикаций, а также тональных портретов различных новостных агрегатов по совокупности освещаемых ими событий в конкретный временной интервал.

Предлагаемый метод, основанный на методах автоматизированного составления словарей по текстам документов, методах семантико-синтаксического анализа и использования онтологических ресурсов, является универсальным поскольку позволяет проводить детальный семантический анализ коллекций документов по различным многомерным критериям. Комплекс тональных словарей большого объема и средства формализации текстов позволят проводить этот анализ с достаточно высокой точностью.

### Литература

1. Пазельская А.Г., Соловьев А.Н. Метод определения эмоций в текстах на русском языке. <http://www.dialog-21.ru/digests/dialog2011/materials/ru/pdf/50.pdf>
2. Клековникова М.В., Котельников Е.В. Метод автоматической классификации текстов по тональности, основанный на словаре эмоциональной лексики. <http://ceur-ws.org/Vol-934/paper15.pdf>
3. Посевкин Р.В., Бессмертный И.А. Применение сентимент анализа текстов для оценки общественного мнения. Научно-технический вестник информационных технологий, механики и оптики, № 1, том 15, 2015.
4. Ермаков С.А., Ермакова Л.М. Методы оценки эмоциональной окраски текстов. Вестник пермского университета. Вып. (1) 9, 2012 .
5. Хорошилов Ал-др. А., Никитин Ю.В., Хорошилов Ал-ей. А., Будско В.И. Автоматическое создание формализованного представления смыслового содержания неструктурированных текстовых сообщений СМИ и социальных сетей // Системы высокой доступности, № 3, том.10, 2014.
6. Калинин Ю.П., Хорошилов Ал-др. А., Хорошилов Ал-ей. А. Современные технологии автоматизированной обработки текстовой информации // Системы высокой доступности, № 2, том.11, 2015.
7. Будско В.И., Калинин Ю.П., Козеренко Е.Б., Хорошилов А.А., Хорошилов А.А. Машинная грамматика русского языка // Системы высокой доступности, № 3, том.13, 2017.
8. Кан А.В., Ревина В.Д., Руснак В.И., Хорошилов Ал-др А., Хорошилов Ал-сей А. Автоматическое формирование синтаксической модели языка для задач машинного перевода и информационного поиска. Сб. «Научно-техническая информация», Сер. 2, № 12, ВИНТИ, 2018.



## Determine the tonality of of news media reports by conceptual analysis

### ***Khoroshilov Al-dr A.***\*

Russian Academy of Sciences, professor of the Moscow  
Aviation Institute (National Research University), Moscow, Russia  
khoroshilov@mail.ru

### ***Kozlovskaya Y.D.***\*\*

Moscow Aviation Institute  
(National Research University), Moscow, Russia  
yana04029877@mail.ru

### ***Musabaev R.R.***\*\*\*

Institute of Information and Computational  
Technologies, Almaty, Kazakhstan  
rmusab@gmail.com

### ***Krasovitskiy A.M.***\*\*\*\*

Institute of Information and Computational  
Technologies, Almaty, Kazakhstan  
akrassovitskiy@gmail.com

### ***Khoroshilov Al-ey A.***\*\*\*\*\*

The 27th Central Research Institute of the Ministry  
of Defence of the Russian Federation, Moscow, Russia  
alex\_khor@mail.ru

The article describes the solution of the task of creating linguistic tools and the methodology of automatic determination of the tonality of news reports related to the quality of life of an ordinary citizen of the Republic of Kazakhstan. The approach

#### **For citation:**

Khoroshilov Al-dr A., Kozlovskaya Y.D., Musabaev R.R., Krasovitskiy A.M., Khoroshilov Al-ey A. Determine the tonality of of news media reports by conceptual analysis. *Modelirovanie i analiz dannykh = Modelling and Data Analysis*, 2019. Vol. 09, no. 4, pp. 67–79. doi: 10.17759/mda.2019090405 (In Russ., abstr. in Engl.)

\****Khoroshilov Aleksander Alekseevich***, Ph.D. (full doctor), lead fellow of the Federal research center «Informatics and management», Russian Academy of Sciences, professor of the Moscow Aviation Institute (National Research University), Moscow. E-mail: khoroshilov@mail.ru

\*\****Kozlovskaya Yana Dmitrievna***, student of the Institute of Moscow Aviation Institute (National Research University), Russia. E-mail: yana04029877@mail.ru

\*\*\****Musabayev Rustam Rafikovich***, Cand.Eng.Sc, Head of the laboratory, Institute of Information and Computational Technologies, Kazakhstan. E-mail: rmusab@gmail.com

\*\*\*\****Krasovitskiy Aleksandr Mikhailovich***, Ph.D., lead fellow of Institute of Information and Computational Technologies, Kazakhstan. E-mail: akrassovitskiy@gmail.com

\*\*\*\*\****Khoroshilov Aleksey Alekseyevich***, Ph.D., Senior Research Officer of The 27th Central Research Institute of the Ministry of Defence of the Russian Federation. E-mail: alex\_khor@mail.ru



for solving the problem was defined, software and methods of automated creation of object dictionaries and evaluation predicate dictionaries, as well as evaluation measure modifier dictionaries were developed. The experiment confirmed the correctness of the proposed methodology for assessing events covered in news reports and the operability of the software complex.

This technique, with appropriate selection of event assessment objects, can be used in creating tonal portraits of specific authors on the set of their publications, as well as tonal portraits of various news aggregates on the set of events they cover in a particular time interval.

**Keywords:** conceptual analysis, predicate-actant structure, tonality of texts, tonal dictionary, tonality modifiers, automatic establishment of assessment of tonality of texts.

### **References**

1. Pazel'skaya A.G., Solov'ev A.N. Metod opredeleniya emotsii v tekstakh na russkom yazyke. <http://www.dialog-21.ru/digests/dialog2011/materials/ru/pdf/50.pdf>
2. Klekovnikova M.V., Kotel'nikov E.V. Metod avtomaticheskoi klassifikatsii tekstov po tonal'nosti, osnovannyi na slovare emotsional'noi leksiki. <http://ceur-ws.org/Vol-934/paper15.pdf>
3. Posevkin R.V., Bessmertnyi I.A. Primenenie sentiment analiza tekstov dlya otsenki obshchestvennogo mneniya. Nauchno-tekhnicheskii vestnik informatsionnykh tekhnologii, mekhaniki i optiki, № 1, tom 15, 2015.
4. Ermakov S.A., Ermakova L.M. Metody otsenki emotsional'noi okraski tekstov. Vestnik permskogo universiteta. Vyp. (1) 9, 2012.
5. Khoroshilov Al-dr. A., Nikitin Yu.V., Khoroshilov Al-ei. A., Budsko V. I. Avtomaticheskoe sozдание formalizovannogo predstavleniya smyslovogo sodержaniya nestrukturirovannykh tekstovykh soobshchenii SMI i sotsial'nykh setei // Sistemy vysokoi dostupnosti, № 3, tom.10, 2014.
6. Kalinin Yu.P., Khoroshilov Al-dr. A., Khoroshilov Al-ei. A. Sovremennye tekhnologii avtomatizirovannoi obrabotki tekstovoi informatsii // Sistemy vysokoi dostupnosti, № 2, tom.11, 2015.
7. Budzko V.I., Kalinin Yu.P., Kozerenko E.B., Khoroshilov A.A., Khoroshilov A.A. Mashinnaya grammatika russkogo yazyka // Sistemy vysokoi dostupnosti, № 3, tom.13, 2017.
8. Kan A.V., Revina V.D., Rusnak V.I., Khoroshilov Al-dr. A., Khoroshilov Al-sei A. Avtomaticheskoe formirovanie sintaksicheskoi modeli yazyka dlya zadach mashinnogo perevoda i informatsionnogo poiska. Sb. «Nauchno-tekhnicheskaya informatsiya», Ser. 2, № 12, VINITI, 2018.