

АКАДЕМИЯ ФСО РОССИИ

На правах рукописи



Толкунов Александр Александрович

**МОДЕЛЬ ОПЕРАТИВНОЙ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ
ТЕКСТОВЫХ КОММЕНТАРИЕВ К ЗАКОНОПРОЕКТАМ**

Специальность 05.13.17 – Теоретические основы информатики

Диссертация на соискание ученой степени кандидата технических наук

Научный руководитель:
Кукушкин Александр Антонович,
кандидат технических наук, доцент

Орёл – 2014

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	4
1 ПРОБЛЕМЫ ОБРАБОТКИ МАССИВА КОММЕНТАРИЕВ, ПОЛУЧЕННОГО В РЕЗУЛЬТАТЕ ОБЩЕСТВЕННОГО ОБСУЖДЕНИЯ ЗАКОНОПРОЕКТОВ.....	9
1.1 Анализ процесса обработки результатов общественного обсуждения законопроектов.....	9
1.1.1 Анализ нормативных документов, регламентирующих процедуру общественного обсуждения.....	9
1.1.2 Анализ процесса обработки комментариев на основе входных и выходных документов.....	15
1.2 Обоснование метода представления и обработки комментариев.	22
1.2.1 Описание формализма понятия "мнение".....	22
1.2.2 Анализ методов обработки мнений, представленных в текстовом виде.....	25
1.2.3 Выбор метода представления и обработки комментариев.	33
1.2.4 Анализ моделей оперативной аналитической обработки текстов.....	35
1.2.5 Концепция системы обработки комментариев.....	51
1.3 Постановка задачи исследования.....	57
Выводы по 1 разделу.....	64
2 РАЗРАБОТКА ЭФФЕКТИВНЫХ ПОДХОДОВ К ОБРАБОТКЕ МАССИВА КОММЕНТАРИЕВ НА ОСНОВЕ ОПЕРАТИВНОЙ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ.....	66
2.1 Модель оперативной аналитической обработки комментариев с операциями детализации на основе выявления типичных и полезных текстов.....	66
2.2 Алгоритм формирования выборки типичных комментариев.....	76
2.3 Алгоритм формирования выборки полезных комментариев.....	84
Выводы по 2 разделу.....	100

3 ОЦЕНИВАНИЕ АДЕКВАТНОСТИ МОДЕЛИ И ЭФФЕКТИВНОСТИ РАЗРАБОТАННЫХ АЛГОРИТМОВ.....	102
3.1 Оценивание и сравнение обобщающих и накапливающих способностей выборок типичных и полезных комментариев.....	103
3.2 Оценивание эффективности алгоритма формирования выборки типичных комментариев.....	109
3.3 Оценивание эффективности алгоритма формирования выборки полезных комментариев.....	115
3.4 Оценивание свойств алгоритма формирования выборки типичных комментариев.....	123
3.5 Оценивание свойств алгоритма формирования выборки полезных комментариев.....	124
Выводы по 3 разделу.....	125
4 НАУЧНО-ТЕХНИЧЕСКИЕ ПРЕДЛОЖЕНИЯ ПО ПОСТРОЕНИЮ И ПРИМЕНЕНИЮ СИСТЕМЫ ОПЕРАТИВНОЙ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ КОММЕНТАРИЕВ.....	127
4.1 Структура системы оперативной аналитической обработки комментариев	127
4.2 Методика применения системы оперативной аналитической обработки комментариев.....	145
Выводы по 4 разделу.....	148
ЗАКЛЮЧЕНИЕ.....	149
СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ.....	151
СЛОВАРЬ ТЕРМИНОВ.....	152
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ.....	154
ПРИЛОЖЕНИЕ А Анализ структуры глав и статей исследуемых законопроектов.....	168
ПРИЛОЖЕНИЕ Б Множества и распределения высказываний в отобранных для исследований классах.....	178
ПРИЛОЖЕНИЕ В Расчет среднего темпа прироста полноты отбора высказываний в выборке типичных комментариев.....	190
ПРИЛОЖЕНИЕ Г Расчет среднего темпа прироста полноты отбора высказываний в выборке полезных комментариев.....	200

ВВЕДЕНИЕ

В настоящее время в соответствии с Постановлением Правительства РФ от 25.08.2012 № 851 [1] разрабатываемые законопроекты должны проходить процедуру общественного обсуждения на интернет-портале regulation.gov.ru [2]. Результатом такой процедуры является массив комментариев, содержащих предложения интернет-пользователей по корректировке законопроекта. Комментарии подлежат экспертной обработке с целью изучения и обобщения поступивших предложений. Трудоемкость данной процедуры на практике оказывается значительной, поскольку число комментариев для законопроектов, вызвавших наибольшее обсуждение, может достигать 10–20 тысяч. В этих условиях возникает потребность в разработке моделей обработки комментариев, направленных на эффективное ознакомление с поступившими предложениями.

Одним из основных понятий при исследовании процедур анализа комментариев является понятие "мнение". Исследованию способов обработки мнений посвящено научное направление Opinion Mining (англ.: анализ мнений), входящее в состав Text Mining (англ.: глубинный анализ текстов). Анализ работ Б. Пэнга, Л. Ли, М. Геймона, Дж. Каренини, М. Девалли, П. Резника, М. Ху, Т. Пинча, Т. Муллена и др. [3–19], посвященных данному направлению исследований, показал, что в основном авторы понимают анализ мнений как задачу классификации по тональности и поиск ключевых слов, в то же время вопросы, касающиеся эффективного ознакомления с мнениями, направленными на улучшение объекта обсуждения, остаются недостаточно исследованными.

Одним из подходов к моделированию процедур обработки мнений может стать подход, учитывающий многомерность представления мнения. Для обработки многомерных данных одним из наиболее эффективных методов обработки является оперативная аналитическая обработка данных, предложенная Э. Ф. Коддом [20].

Модификации данного метода обработки для текстов описаны в трудах Дж. Мозе, С. Кейта, Х. Ло, С. Лин, Дж. Хана, Д. Жана, Й. Ю, Б. Джанета, А. Иноокучи, К. Такеда и др. [21–31] Однако анализ данных работ показывает, что предлагаемые модели не универсальны, авторы вводят иерархии категорий, актуальные в конкретной задаче, а вновь вводимые операции погружения ограничивают лишь списками ключевых слов.

В данной ситуации возникли условия для постановки и решения *научной задачи* по разработке модели эффективной обработки комментариев с использованием оперативной аналитической обработки текстов. Разработанная модель обработки комментариев обеспечивает поддержку рекурсивной процедуры понимания комментариев экспертом за счет последовательного перехода между классами комментариев с погружением и ознакомлением с предложениями по корректировке законопроекта по дедуктивной схеме: интенсивность мнений в классе комментариев – ознакомление с типичными комментариями – ознакомление с полезными комментариями – ознакомление со всеми комментариями.

Объектом исследования является процесс аналитической обработки комментариев интернет-пользователей в ходе общественного обсуждения законопроектов.

В качестве *предмета исследования* выступают модели оперативной аналитической обработки текстов и алгоритмы погружения в текстовые данные в ходе обработки результатов общественного обсуждения законопроектов.

Целью исследования является повышение темпа изучения совокупности высказываний в ходе обработки результатов общественного обсуждения законопроектов.

Методы исследований базируются на использовании аппарата математической статистики, теории оптимизации, моделей представления знаний, моделей семантической фильтрации текстов, методов классификации, кластеризации и реферирования текстов.

Научная новизна работы обусловлена следующим:

1. Для решения задачи построения модели гиперкуба использованы зависимости между атрибутами комментариев.

2. Для решения задачи выделения типичных комментариев в ячейках гиперкуба применены кластеризация на основе матриц подобия и принцип наибольшего содержательного подобия со сводным рефератом кластера.

3. Решение задачи отбора полезных комментариев в ячейках гиперкуба осуществлено на основе максимизации функционала обобщенной полезности комментария, сформированного с учетом выявленной аналитической зависимости свертки атрибутов конструктивных предложений, составляющих комментарий, и предпочтений пользователей от ранга комментария.

Теоретическая ценность диссертации заключается в разработке решений, направленных на развитие моделей представления и обработки знаний в форме гиперкуба для обработки текстовой информации на основе операций определения типичных и полезных текстов.

Практическая значимость исследования определяется возможностью повышения оперативности анализа мнений интернет-пользователей по обсуждаемым вопросам на основе технической реализации и применения разработанных модели и алгоритмов в деятельности информационно-аналитических и ситуационных центров государственных органов.

Реализация результатов исследования осуществлена в практику деятельности ведомственных аналитических подразделений и в учебный процесс Академии ФСО России, что подтверждается соответствующими актами.

Содержание диссертации соответствует паспорту специальности 05.13.17 – Теоретические основы информатики (технические науки) по следующим областям исследований:

п. 1. Исследование, в том числе с помощью средств вычислительной техники, информационных процессов, информационных потребностей коллективных и индивидуальных пользователей;

п. 2. Исследование информационных структур, разработка и анализ моделей информационных процессов и структур;

п. 5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

Научные результаты, выносимые на защиту:

1. Модель оперативной аналитической обработки комментариев, отличающаяся от известных способом формирования гиперкуба на основе многомерного представления мнения, а также набором операций погружения в текстовые данные, учитывающих неравномерность распределения высказываний в комментариях интернет-пользователей.

2. Комплекс алгоритмов, включающий алгоритм формирования выборки типичных комментариев, учитывающий содержательное подобие комментариев и сходство с типичным представителем мнения для приоритетного отбора комментариев, и алгоритм формирования выборки полезных комментариев, учитывающий атрибуты конструктивных предложений и предпочтения интернет-пользователей для приоритетного отбора комментариев.

3. Научно-технические предложения по построению и применению системы оперативной аналитической обработки комментариев, позволяющие производить их обработку с учетом неравномерного распределения высказываний и сходства с типичным представителем мнения при заданном бюджете времени.

Достоверность выводов и рекомендаций обусловлена корректностью применения методов математической статистики, теории оптимизации, методов обработки текстов, воспроизводимостью и проверяемостью теоретических и экспериментальных результатов, согласованностью с практикой, внутренней непротиворечивостью, практической реализацией полученных результатов при разработке автоматизированных способов анализа комментариев.

Личный вклад соискателя. Все изложенные в диссертации результаты исследования получены соискателем лично с учетом замечаний и рекомендаций научного руководителя.

Во введении обоснованы актуальность исследования, его научная новизна. Определены объект, предмет, цель исследования. Перечислены научные результаты, выносимые на защиту, приведены сведения о публикациях, апробациях и реализации результатов.

В первом разделе проведен анализ процесса обработки комментариев, произведен анализ решений, позволяющих повысить эффективность данного процесса с учетом многомерности мнений, и произведена постановка задачи.

Во втором разделе приведено описание модели оперативной аналитической обработки комментариев и алгоритмов формирования выборок типичных и полезных комментариев.

В третьем разделе произведено оценивание адекватности модели и эффективности разработанных алгоритмов.

В четвертом разделе описаны структура и методика применения системы оперативной аналитической обработки комментариев.

В заключении сформулированы выводы по результатам проведенного исследования.

Результаты исследования обсуждались на Международной молодежной научно-практической конференции "ИНФОКОМ-2013" (г. Ростов-на-Дону, Северо-Кавказский филиал Московского технического университета связи и информатики, 2013), VIII Всероссийской межведомственной научной конференции "Актуальные проблемы развития технологических систем государственной охраны, специальной связи и специального информационного обеспечения" (г. Орёл, Академия ФСО России, 2013), V Межвузовской научно-практической конференции "Проблемы и перспективы совершенствования охраны Государственной границы" (г. Калининград, Калининградский пограничный институт ФСБ России, 2012), VII Научно-практической конференции "Проблемы развития технологических систем государственной охраны, специальной связи и специального информационного обеспечения" (г. Орёл, Академия ФСО России, 2011).

По теме диссертации опубликовано 8 печатных работ, из них 3 – в журналах из Перечня ВАК Минобрнауки РФ [80, 94, 105], получено 2 свидетельства о регистрации программ для ЭВМ в Роспатенте.

1 ПРОБЛЕМЫ ОБРАБОТКИ МАССИВА КОММЕНТАРИЕВ, ПОЛУЧЕННОГО В РЕЗУЛЬТАТЕ ОБЩЕСТВЕННОГО ОБСУЖДЕНИЯ ЗАКОНОПРОЕКТОВ

1.1 Анализ процесса обработки результатов общественного обсуждения законопроектов

1.1.1 Анализ нормативных документов, регламентирующих процедуру общественного обсуждения

В настоящее время в Российской Федерации действует порядок подготовки нормативных правовых актов, обязательным этапом которого является процедура общественного обсуждения разрабатываемых документов.

Регулирование данной процедуры осуществляется следующими нормативными правовыми актами:

1. Указ Президента РФ от 9 февраля 2011 г. № 167 "Об общественном обсуждении проектов федеральных конституционных законов и федеральных законов" [32].

Документ устанавливает необходимость вынесения на общественное обсуждение проектов федеральных законов по решению Президента Российской Федерации. Обсуждаемые законопроекты и информация о порядке направления гражданами замечаний и предложений должны быть размещены в сети Интернет на официальных или специально созданных сайтах [2, 33] с указанием времени, в течение которого планируется общественное обсуждение.

2. Постановление Правительства РФ от 22.02.2012 № 159 "Об утверждении Правил проведения общественного обсуждения проектов федеральных конституционных законов и федеральных законов" [34].

Документ устанавливает порядок проведения общественного обсуждения законопроектов с использованием сети Интернет, включая перечень документов, прилагаемых к законопроекту, выбор экспертов, выбор организа-

ции-оператора, сроки обсуждения, порядок представления отчета по результатам анализа поступивших замечаний и предложений, вопросы по рассмотрению доработанных законопроектов.

3. Постановление Правительства РФ от 25 августа 2012 г. № 851 "О порядке раскрытия федеральными органами исполнительной власти информации о подготовке проектов нормативных правовых актов и результатах их общественного обсуждения" [1].

Документ определяет, что с 15 апреля 2013 г. вся информация о подготовке проектов нормативных правовых актов, а также результатах их общественного обсуждения размещается на официальном сайте regulation.gov.ru [2] (рисунок 1.1).

Рисунок 1.1 – Главная страница сайта regulation.gov.ru

В Постановлении подробно описаны действия разработчика нормативно-правового акта по размещению на официальном сайте информации о подготовке, сроках и результатах обсуждения и рассмотрения разрабатываемого документа.

4. Постановление Правительства РФ от 13 августа 1997 г. № 1009 "Об утверждении правил подготовки нормативных правовых актов федеральных органов исполнительной власти и их государственной регистрации" (в ред. Постановления Правительства РФ от 18.12.2012 № 1334) [35].

Документ требует при представлении на государственную регистрацию нормативного правового акта прилагать к нему справку, содержащую сведения о раскрытии информации по разработке документа в соответствии с Правилами, утвержденными Постановлением Правительства РФ от 25 августа 2012 г. № 851.

5. Постановление Правительства РФ от 2 мая 2012 г. № 421 "О мерах по совершенствованию подготовки нормативных правовых актов федеральных органов исполнительной власти, устанавливающих не относящиеся к сфере технического регулирования обязательные требования" [36].

Документ устанавливает перечень нормативных правовых актов, в отношении которых проводится публичное обсуждение, и уточняет порядок их разработки и рассмотрения.

6. Постановление Правительства РФ от 26 февраля 2012 г. № 96 "Об антикоррупционной экспертизе нормативных правовых актов и проектов нормативных правовых актов" [37].

Документ определяет, что одним из этапов проведения независимой антикоррупционной экспертизы проектов нормативных правовых актов является их размещение и общественное обсуждение на официальном сайте в сети Интернет.

7. Постановление Правительства РФ от 29 июля 2011 г. № 633 "Об экспертизе нормативных правовых актов федеральных органов исполнительной

власти в целях выявления в них положений, необоснованно затрудняющих ведение предпринимательской и инвестиционной деятельности" [38].

Документ указывает, что в целях выявления в нормативных правовых актах положений, необоснованно затрудняющих ведение инвестиционной и предпринимательской деятельности, при необходимости проводится общественная экспертиза.

8. Постановление Правительства Российской Федерации от 16 мая 2011 г. № 373 "О разработке и утверждении административных регламентов исполнения государственных функций и административных регламентов предоставления государственных услуг" [39].

Документ определяет, что проекты регламентов размещаются на официальном сайте в сети Интернет для их общественного обсуждения.

9. Постановление Правительства Российской Федерации от 17 декабря 2012 г. № 1318 "О порядке проведения федеральными органами исполнительной власти оценки регулирующего воздействия проектов нормативных правовых актов, проектов поправок к проектам федеральных законов и проектов решений Совета Евразийской экономической комиссии, а также о внесении изменений в некоторые акты Правительства Российской Федерации" [40].

Документ устанавливает, что информация о разработке федеральными органами исполнительной власти проектов актов, а также о результатах их публичного обсуждения размещается на официальном сайте regulation.gov.ru в сети Интернет.

10. Федеральный закон от 27 июля 2010 г. № 210-ФЗ "Об организации предоставления государственных и муниципальных услуг" [41].

В документе содержатся требования о том, чтобы проекты разрабатываемых административных регламентов размещались на официальном сайте в сети Интернет для прохождения независимой экспертизы.

11. Приказ Минэкономразвития России от 27 мая 2013 г. № 290 "Об утверждении формы сводного отчета о проведении оценки регулирующего

воздействия, формы заключения об оценке регулирующего воздействия, методики оценки регулирующего воздействия" [42].

Документ устанавливает перечень форм сводного отчета о проведении оценки регулирующего воздействия проекта акта, включающих сведения о сроках публичного обсуждения проекта акта, лицах, представивших предложения, структурных подразделениях разработчика, рассмотревших представленные предложения. Приложением к сводному отчету является сводка предложений с указанием сведений об их учете либо причинах отклонения.

На основе анализа действующих нормативных документов можно выделить следующие основные этапы подготовки нормативных правовых актов с учетом требований по проведению их общественного обсуждения (рисунок 1.2):

1. Уведомление о подготовке проекта нормативного правового акта.

Уведомление размещается разработчиком на официальном сайте в сети Интернет. Срок его общественного обсуждения определяется разработчиком и составляет не менее 15 календарных дней.

2. Разработка проекта нормативного правового акта и его общественное обсуждение.

Срок общественного обсуждения проекта нормативного правового акта составляет от 15 до 60 календарных дней в зависимости от типа документа. Обсуждение проводится на сайте regulation.gov.ru.

3. Составление сводного отчета.

Сводный отчет выполняется по результатам рассмотрения всех поступивших предложений и размещается на официальном сайте не позднее дня направления нормативного правового акта на согласование с ФОИВ.

4. Подготовка заключения об оценке регулирующего воздействия.

Заключение составляется на основе установленных форм в Минэкономразвития РФ.

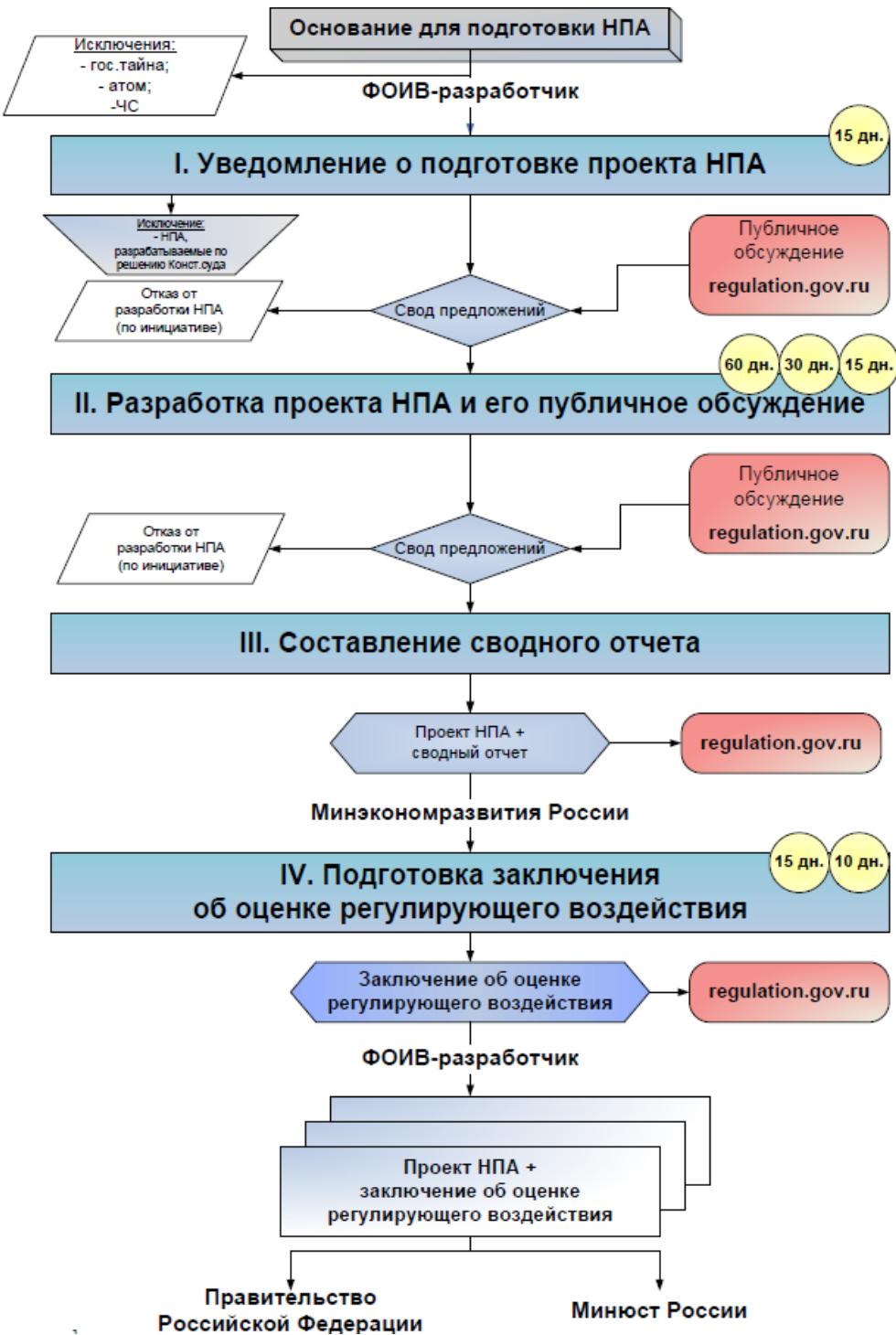


Рисунок 1.2 – Этапы разработки нормативных правовых актов с учетом требований по проведению общественного обсуждения

Действующее законодательство устанавливает обязательные требования по проведению и обработке результатов общественного обсуждения в заданные сроки, однако не определяет порядок обработки поступивших предложений.

1.1.2 Анализ процесса обработки комментариев на основе входных и выходных документов

В ходе общественного обсуждения интернет-пользователи знакомятся с содержанием законопроекта и выражают свое мнение по содержанию отдельных статей в форме текстовых комментариев в специально предназначенному для этого разделе официального сайта. Практика данной процедуры на примере законопроектов "О полиции" и "Об образовании" показывает, что результат данной процедуры представляет собой массив мнений объемом до 10–20 тыс. текстовых комментариев.

Текст комментария имеет типовую структуру (рисунок 1.3):

Tatyana, 15 сентября, 11:59 # // Начало комментария 1
 К главе 2, статье 6 // Обсуждаемая статья
 Прописать конкретно неуместные ситуации // Содержание комментария 1
 (при вооруженном сопротивлении, проведении контртеррористических операций и т. п.), иначе – широкие трактовки. Либо вообще убрать неуместно.
 +0 // Конец комментария 1

Сергей, 15 сентября, 11:54 # // Начало комментария 2
 К главе 5, статье 21 // Обсуждаемая статья
 В качестве специального средства // Содержание комментария 2
 полицейский также должен иметь право использовать служебный автомобиль, а именно для задержания лица, застигнутого при совершении преступления и пытающегося скрыться, и в случае невыполнения лицом, управляющим транспортным средством, требования сотрудника полиции об остановке. При помощи служебного ТС возможно преграждение пути следования правонарушителя.
 +0 // Конец комментария 2

Рисунок 1.3 – Пример комментария

Комментарий включает следующие элементы:

- автор комментария;
- дата и время составления комментария;
- глава и статья законопроекта, к которой относится комментарий;
- текст комментария (в соответствии с "Правилами проведения общественного обсуждения ..." объем комментария ограничен 4000 (четырьмя тысячами) знаков);

– число "лайков" – голосов респондентов в поддержку мнения, изложенного в тексте комментария.

Все тексты комментариев, полученных по результатам общественного обсуждения одного законопроекта, доступны для просмотра на одной динамически формируемой странице официального сайта в разделе "Архив обсуждения". Имеется возможность просмотра комментариев, относящихся к отдельным статьям законопроекта. Комментарии доступны для копирования. При сплошном копировании комментариев в текстовый файл их тексты располагаются в нем последовательно, при этом текст каждого комментария соответствует типовой структуре.

Порядок обработки результатов общественного обсуждения детально не регламентирован. Для подготовки аналитического обзора (доклада) по результатам обработки поступивших предложений на практике создают комиссию из 10–15 человек, в состав которой входят представители профильных министерств, ведомств, эксперты в области обсуждаемой темы.

Аналитический обзор (доклад) также доступен для просмотра и копирования на одной динамически формируемой странице официального сайта в разделе "О законопроекте". Материал аналитического обзора (доклада) удобен для работы при представлении его в формате MS Word ".doc".

Структура текста доклада нормативно не определена. Анализ логической структуры текста доклада показывает, что он структурируется на 2 блока:

- блок, содержащий описание статистики массива комментариев к законопроекту (вводная часть доклада);
- блок, содержащий постатейный анализ с выделением типичных мнений и примерами полезных комментариев (выдержек комментариев), отражающих данные мнения (основная часть доклада).

Анализ содержания аналитического обзора показывает, что действия эксперта по изучению и пониманию поступивших предложений представлены следующей последовательностью этапов (рисунок 1.4):

- классификация мнений интернет-пользователей,
- выявление и ранжирование на основе интенсивности мнений по классам "болевых точек" в тексте законопроекта,
- последовательное ознакомление с содержанием мнений в классах,
- формирование типичных представителей мнений по классам,
- формирование выборки полезных комментариев, подтверждающих данные типичные мнения:



Рисунок 1.4 – Последовательность этапов обработки результатов общественного обсуждения экспертом

При проведении анализа массива комментариев в качестве классификационных признаков для группировки комментариев эксперты используют деление комментариев по главам и статьям. В то же время в ходе анализа комментариев исходного массива и обзора установлено, что в качестве дополнительных признаков для деления комментариев на классы, позволяющих более детально типизировать и изучить поступившие предложения, следует использовать следующие классификационные признаки:

- функциональный стиль комментария,
- оценочная тональность комментария,
- эмоциональная тональность комментария.

Указанные признаки определяют и позволяют выделить крупные устойчивые классы комментариев в исходном массиве и обзоре (рисунки 1.5–1.7).

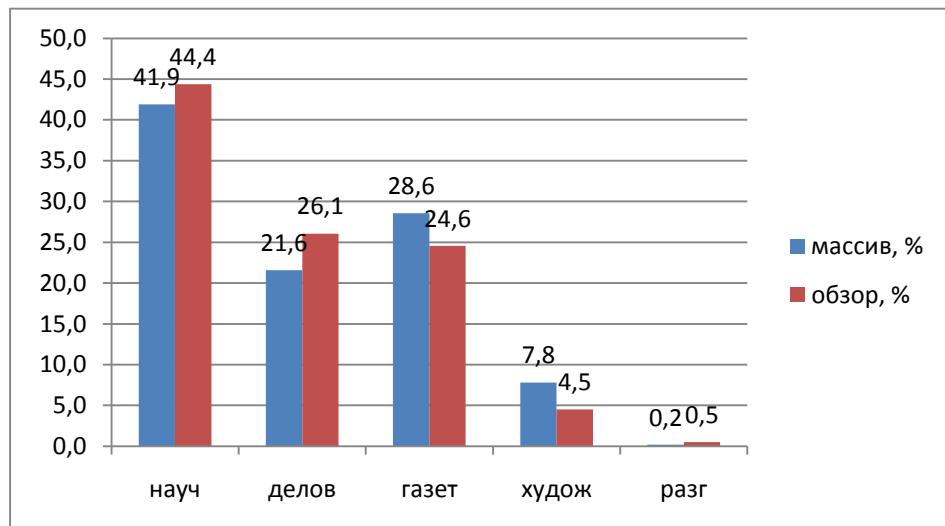


Рисунок 1.5 – Доли комментариев различных функциональных стилей в исходном массиве и обзоре к законопроекту "О полиции"

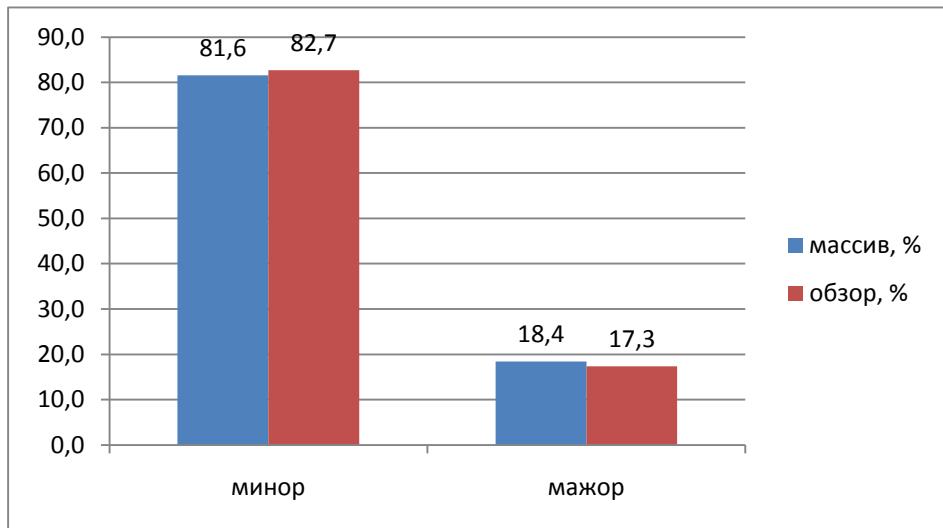


Рисунок 1.6 – Доли комментариев различной оценочной тональности в исходном массиве и обзоре к законопроекту "О полиции"

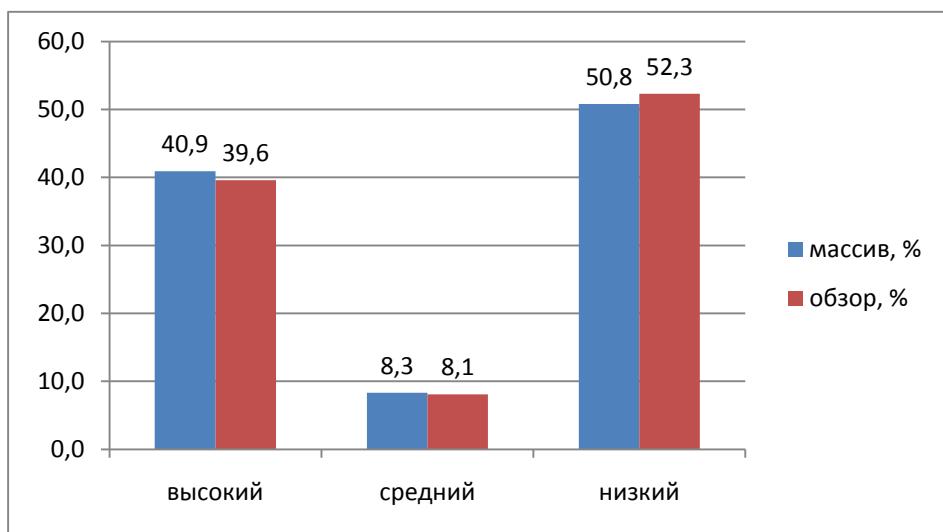


Рисунок 1.7 – Доли комментариев различной эмоциональной тональности в исходном массиве и обзоре к законопроекту "О полиции"

Анализ порядка обращения к статьям законопроекта и содержания отобранных комментариев в аналитическом обзоре позволил выявить следующие закономерности проведения эвристических процедур экспертами:

- значительное влияние интенсивности мнения E , выраженной числом комментариев, на очередность и глубину анализа статей;

- стремление подкреплять типичное мнение по классам комментариями более строгих стилей (официально-деловой, научный);
- наличие в отобранных экспертами комментариях конструктивных идей и предложений.

Помимо этого анализ структуры и содержания аналитического обзора показывает, что при его формировании эксперты учитывают следующие формальные признаки мнений (таблица 1.1).

Таблица 1.1 – Учет формальных признаков мнений при формировании аналитического обзора

Эвристическая процедура	Раздел доклада	Формальные составляющие мнений
Выявление "болевых точек"	Вводная часть	Глава/Статья (Q); Автор($\{A_i\}$); Дата/Время (T); Интенсивность (E)
Выявление типичных мнений по статьям	Основная часть (анализ мнений по статьям)	Глава/Статья (Q); Тональность (V)
Отбор комментариев по статьям	Основная часть (анализ мнений по статьям)	Глава/Статья (Q); Стиль (S); Тональность (V); Полезность (U)

Значительное влияние на процедуру обработки массива комментариев экспертами оказывает его объем. Для оценивания среднего времени, необходимого для простого прочтения массива комментариев к закону одним экспертом, целесообразно использовать выражение:

$$T = N_c \cdot \bar{N}_w \cdot \bar{t}_{1w}, \quad (1.1)$$

где N_c – количество комментариев, \bar{N}_w – среднее количество слов в комментарии, \bar{t}_{1w} – среднее время чтения одного слова. Так, имея для закона "О полиции" следующие значения параметров:

$$N_c = 20\,000 \text{ комментариев}; \quad (1.2)$$

$$\bar{N}_w = 73,5 \text{ слов/комментарий}; \quad (1.3)$$

$$\bar{t}_{1w} = 0,3 \text{ с,} \quad (1.4)$$

получаем, что трудоемкость простого прочтения массива комментариев одним экспертом составляет:

$$T = 20\,000 \cdot 73,5 \cdot 0,3 = 441\,000 \text{ (с)} \approx 123 \text{ (часа)} \quad (1.5)$$

При 8-часовом рабочем дне одному эксперту для решения данной задачи (непрерывного чтения) потребуется 16 рабочих дней. Указанные затруднения частично снимаются назначением в комиссию по рассмотрению комментариев, поступивших в ходе общественного обсуждения, группы экспертов по теме обсуждения. При этом необходимо отметить, что групповая работа по анализу поступивших комментариев требует временных ресурсов на согласование знаний экспертов о предложениях по корректировке обсуждаемого законопроекта. Снизить трудозатраты на обработку комментариев возможно путем внедрения в процесс работы эксперта автоматизированных процедур.

По результатам анализа процесса обработки комментариев на основе входных и выходных документов можно сделать следующие выводы:

- процесс обработки комментариев направлен на ознакомление с предложениями по корректировке разрабатываемого законопроекта путем деления исходного массива комментариев на классы, выявление в полученных классах типичных мнений и подкрепление их полезными комментариями,
- деление комментариев по главам и статьям формирует классы с большим числом комментариев и может быть дополнительно детализировано с использованием признаков стиля и тональности,
- пониманию типичных и полезных мнений в классах предшествует процедура последовательного изучения комментариев, требующая значительных трудозатрат.

1.2 Обоснование метода представления и обработки комментариев

1.2.1 Описание формализма понятия "мнение"

Важным при обработке массива комментариев является понятие "мнение". Подход к формализации указанного понятия предпринимался еще в трудах Г. Ф. Гегеля [43], который выделял во мнении ряд структурных элементов:

- условие существования;
- объект (содержание);
- носитель;
- характер суждения;
- соотношение "всеобщего" и "особыенного" мнений.

Выделим формальные элементы мнений на основе толкования понятия "мнение" в современных источниках [44–48] (таблица 1.2):

Таблица 1.2 – Формальные элементы понятия "мнение" на основе вариантов толкования в современных источниках

Источник	Понятие	Формальные элементы
С. И. Ожегов, Н. Ю. Шведова Толковый словарь русского языка	Суждение, выражающее оценку чего-нибудь, отношение к кому-чему-нибудь, взгляд на что-нибудь	W – текст высказывания; F – факт, относительно которого высказывают мнение; V – оценка факта; O – отношение к факту; $\{A_i\}$ – автор(ы) мнения
Словарь Т. Ф. Ефремовой	Суждение, выражающее оценку кого-либо, чего-либо, отношение к кому-либо, чему-либо, взгляд на кого-либо, что-либо. Официальное заключение по какому-нибудь вопросу, требующему решения. Оценка, то или иное суждение о ценности чего-нибудь	W – текст высказывания; F – факт, относительно которого высказывают мнение; V – оценка факта; O – отношение к факту; $\{A_i\}$ – автор(ы) мнения

Продолжение таблицы 1.2

Источник	Понятие	Формальные элементы
Словарь Д. Н. Ушакова	Взгляд на что-нибудь, суждение о чем-нибудь, выраженное в словах	W – текст высказывания; F – факт, относительно которого высказывают мнение; V – оценка факта; $\{A_i\}$ – автор(ы) мнения
ru.wikipedia.org	Форма сознания, в котором проявляется отношение (скрытое или явное) различных людей к событиям и процессам действительной жизни, затрагивающим их интересы и потребности	F – факт, относительно которого высказывают мнение; O – отношение к факту; $\{A_i\}$ – автор(ы) мнения; Q – аспект факта, по которому высказывают мнение
Современный толковый сло- варь издатель- ства "Большая Советская Эн- циклопедия"	Состояние сознания, заключающее в себе отношение (скрытое или явное) к общественным событиям, к деятельности различных групп, организаций, отдельных личностей; выражает позицию одобрения или осуждения по тем или иным общественным проблемам, регулирует поведение индивидов, социальных групп и институтов, насижает определенные нормы общественных отношений; действует как в рамках общества в целом, так и в рамках различных социальных групп. Активность и значение мнения определяются характером социальной структуры общества, уровнем развития экономической культуры, демократических институтов и свобод	F – факт, относительно которого высказывают мнение; O – отношение к факту; $\{A_i\}$ – автор(ы) мнения; D – степень доверия мнению (обоснованность мнения); C – цель высказывания мнения; S – оценка стиля высказывания; T – время; Q – аспект факта, по которому высказывают мнение; E – интенсивность проявления мнения; U – оценка полезности мнения

Анализ источников, характеризующих понятие "мнение", позволяет ввести для него теоретико-множественное представление в виде кортежа:

$$M = \langle W, F, V, O, \{A_i\}, D, C, S, T, Q, E, U \rangle, \quad (1.6)$$

где W – текст высказывания; F – факт, относительно которого высказывают мнение; V – оценка факта; O – отношение к факту; $\{A_i\}$ – автор(ы) мнения;

D – степень доверия мнению (обоснованность мнения); C – цель высказывания мнения; S – оценка стиля высказывания; T – время; Q – аспект факта, по которому высказывают мнение; E – интенсивность проявления мнения; U – оценка полезности мнения.

Данное теоретико-множественное представление является общим и требует уточнения состава формальных элементов применительно к решаемой задаче обработки результатов общественного обсуждения законопроектов. С учетом формальных признаков мнений, применяемых при формировании содержания аналитического обзора экспертами (таблица 1.1), мнение в отношении законопроекта в эвристиках экспертов принимает вид:

$$M_{zp} = \langle W, V, \{A_i\}, S, T, Q, E, U \rangle. \quad (1.7)$$

Полученный формализм позволяет осуществлять выбор метода представления и обработки комментариев, полученных в результате общественного обсуждения законопроектов, за счет:

- определения способа классификации и группировки исходного массива комментариев по различным значимым группам признаков;
- определения порядка обращения к классам комментариев с использованием интенсивности мнений (например, количества комментариев) в классах;
- определения порядка изучения комментариев в классах путем установления сходства и различия высказанных предложений по корректировке законопроекта;
- определения порядка изучения комментариев в классах путем установления полезности комментария и ранжирования комментариев по данному показателю.

1.2.2 Анализ методов обработки мнений, представленных в текстовом виде

Задачи обработки мнений, представленных в текстовом виде, развиваются в рамках направления Opinion Mining (англ.: анализ мнений), входящего в состав Text Mining (англ.: глубинный анализ текстов). В рамках данного направления предложены методы обработки, основанные на методах автоматической обработки текстов. К основным методам обработки мнений следует отнести:

- классификацию/ранжирование по оценочной и эмоциональной тональностям;
- классификацию по степени высказанной или полученной мнением поддержки;
- классификацию мнений по смысловой противоположности;
- извлечение ключевых слов и понятий.

В [3–19] описаны следующие примеры применения данных методов для обработки коллекций документов:

1. Построение графовых моделей для описанных в тексте субъектов и мнениях данных субъектов друг о друге.

Пример графовой модели мнений субъектов приведен на рисунке 1.8. Указанная модель представляет результат ручной обработки массива мнений, связанных с действиями различных должностных лиц и организаций по предупреждению и устраниению последствий урагана "Катрина" в США [4]. Узлами графа являются субъекты, являющиеся источником или предметом высказываний. Граф дополнен выдержками текста высказываний, связанными с соответствующими дугами и характеризующими негативную тональность мнений.

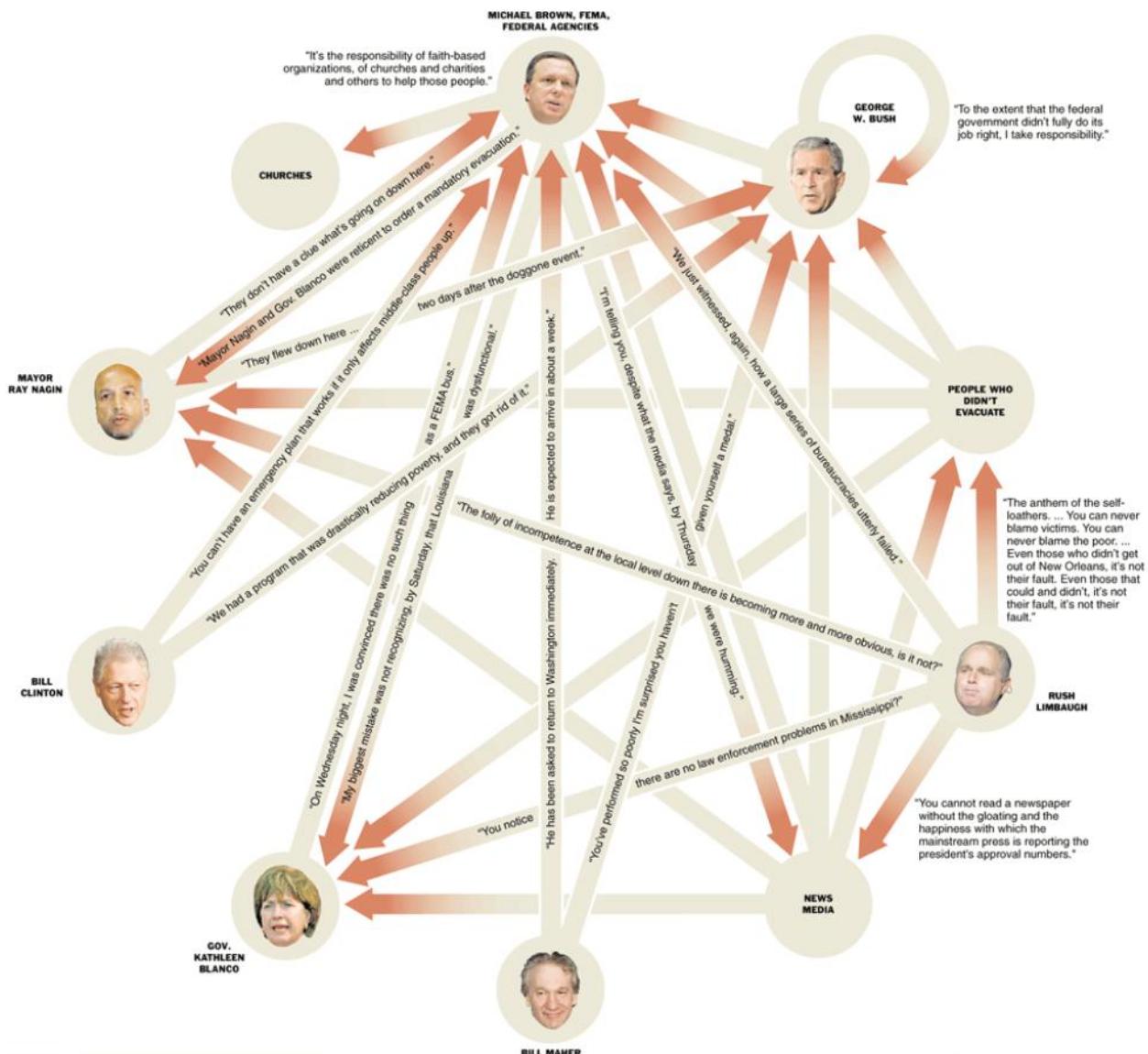


Рисунок 1.8 – Графовая модель мнений субъектов

Достоинством подобного подхода является наглядность представления взаимодействий небольшого числа субъектов в ходе высказывания мнений.

Недостатком подхода является значительный рост числа узлов и ребер графа при увеличении числа субъектов. Кроме того, построение данного типа моделей в ручном режиме снижает их потенциал как средства поддержки принятия решений, поскольку эксперт предварительно знакомится с массивом мнений и на момент построения графа уже понимает содержание высказываний, а значит, в модели нет необходимости.

Применение автоматического режима построения графа мнений [5] частично устраняет подобный недостаток (рисунок 1.9). В этом случае источ-

ники и предмет их высказываний, а также тональность и степень ее проявления извлекаются из текста автоматически и применяются для построения графа.

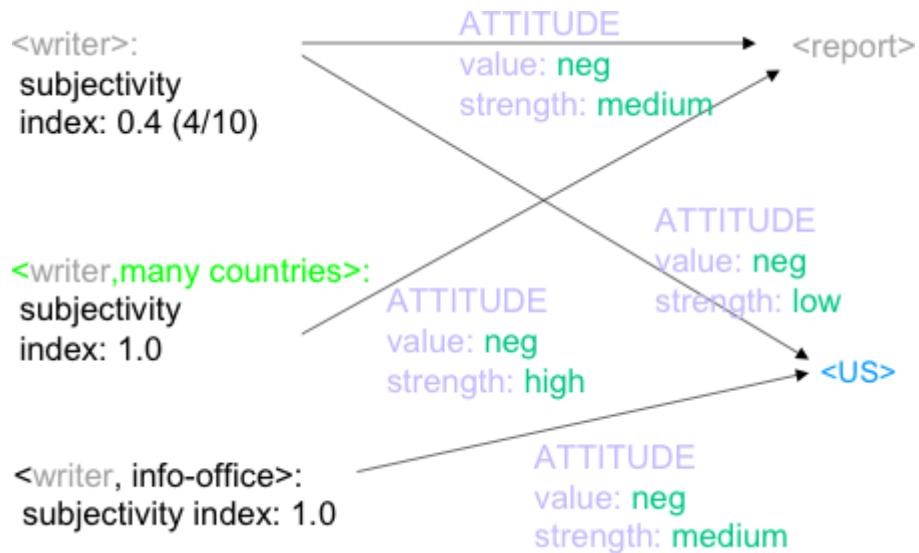


Рисунок 1.9 – Графовая модель мнений, построенная в автоматическом режиме

При построении графа мнений в автоматическом режиме сложность его интерпретации высока, поскольку в различных их реализациях авторы применяют собственные обозначения, символы и графические элементы.

2. Построение обзорных рефератов.

Для построения рефератов могут применяться как обычные, так и модифицированные алгоритмы реферирования, учитывающие особенности мнений.

Простым способом адаптации имеющихся алгоритмов является модификация входных данных, например формирование подмассива документов, связанных общей темой либо принадлежащих одному стилю [6].

Модификацией данного способа является учет дополнительно к тематическому разбиению мнений предварительно определенной иерархии характеристик обсуждаемого продукта [7]. Такой подход позволяет получить более детализированные обобщения (рисунок 1.10).

Summary created via a "true natural-language-generation" approach:

Almost all users loved the Canon G3 possibly because some users thought the physical appearance was very good. Furthermore, several users found the manual features and the special features to be very good. Also, some users liked the convenience because some users thought the battery was excellent. Finally, some users found the editing/viewing interface to be good despite the fact that several customers really disliked the viewfinder. However, there were some negative evaluations. Some customers thought the lens was poor even though some customers found the optical zoom capability to be excellent. Most customers thought the quality of the images was very good.

Summary created by a modified sentence-extraction system:

Bottom line, well made camera, easy to use, very flexible and powerful features to include the ability to use external flash and lens/filters choices. It has a beautiful design, lots of features, very easy to use, very configurable and customizable, and the battery duration is amazing! Great colors, pictures, and white balance. The camera is a dream to operate in automode, but also gives tremendous flexibility in aperture priority, shutter priority, and manual modes. I'd highly recommend this camera for anyone who is looking for excellent quality pictures and a combination of ease of use and the flexibility to get advanced with many options to adjust if you like.

Рисунок 1.10 – Сравнение обобщенного (сверху) и детализированного рефератов

Помимо различных разбиений исходного массива известны и способы построения обзорных рефератов, рассматривающие тексты массива как комбинацию отдельных предложений. Один из таких подходов [8] для текстов, объединенных общей темой, основан на оценке тональной полярности входящих в тексты предложений по содержащимся в них тонально окрашенным словам и отрицаниям. Для получения обзорного реферата документов с положительной либо отрицательной тональностью определяется документ, содержащий наиболее полярные предложения соответствующей тональности, и в качестве реферата выступает его заголовок.

Нестандартным подходом к "реферированию" массива мнений считается подход, основанный на выборе отдельных документов для представления эксперту.

В этом направлении описан подход [9], в рамках которого для выбранного экспертом текста производится оценивание тональной окраски лексики, по результатам которого из массива выбирается текст, содержащий лексику аналогичной либо противоположной окраски, и рекомендуется для изучения эксперту.

В обзоре методов обработки web-данных [10] отмечена необходимость разработки методов анализа массивов мнений, основанных на формировании выборок, учитывающих наличие разных мнений, а не только принадлежащих доминирующей тональности.

3. Статистическая обработка мнений с последующей визуализацией результатов.

Примером такой обработки является подсчет долей положительных/нейтральных/отрицательных мнений об обсуждаемом объекте с последующей визуализацией в форме "термометра", шкала которого поделена в соответствующей пропорции на отрезки с использованием цветографической разметки. Также известен вариант графической интерпретации статистики мнений на основе количества звезд, характеризующих уровень высказанной в мнении поддержки обсуждаемого объекта.

Альтернативным подходом к представлению результата статистической обработки мнений является метод, описанный в [11]. В данном подходе вся совокупность мнений изображается в виде некоторой геометрической фигуры, площадь которой разбивается на сектора в соответствии с распределением комментариев по обсуждаемым в них характеристикам объекта. К полученным секторам применяется цветографическая разметка, характеризующая обобщенную тональность высказанных по данной характеристике объекта мнений (рисунок 1.11).

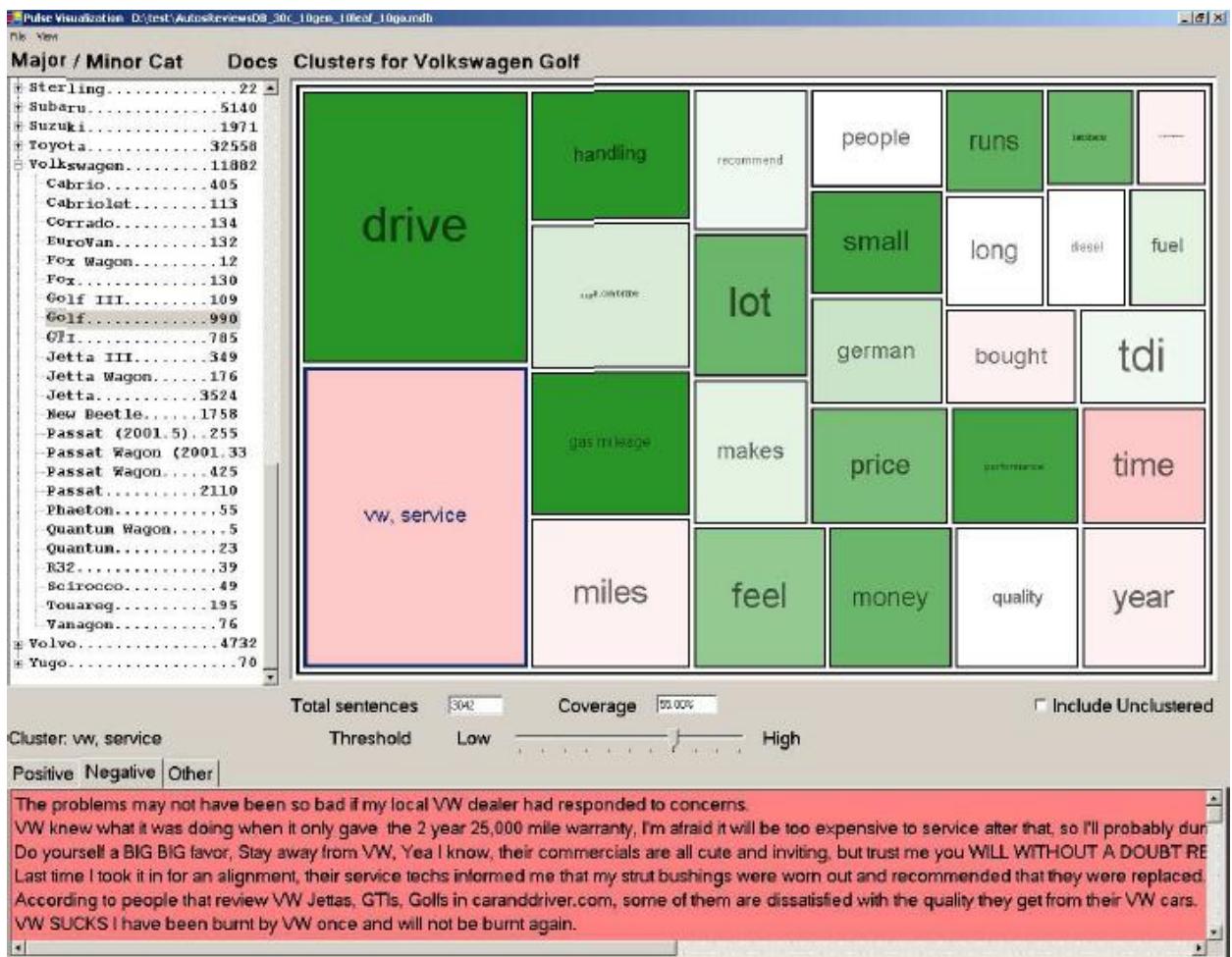


Рисунок 1.11 – Геометрическая интерпретация статистики мнений с цветографической разметкой по тональности

В [12] для представления статистики мнений применяются столбцы, характеризующие свойства обсуждаемого объекта, и линии горизонта. Размер столбца характеризует количество мнений, относящихся к некоторому свойству объекта, а его положение относительно линии горизонта – соотношение мнений различной тональности в отношении данного свойства. Разновидностью данного вида представлений являются графики, на которых столбцы выравниваются не по линии горизонта, а по основанию. Указанные графики более удобны для сравнения интенсивности мнений в отношении различных свойств объекта, но затрудняют сравнение интенсивностей мнений определенной тональности.

Альтернативой методам представления статистики мнений с использованием столбцов являются методы на основе лепестковых диаграмм, в кото-

рых каждый лепесток характеризует заданное свойство объекта [13]. Помимо лепестков, связанных с положительной/отрицательной тональностью мнений, в диаграмму включаются лепестки, характеризующие свойства обсуждаемого объекта.

Для определения связи между обсуждаемыми объектами и характеризующими их терминами предопределенной полярности в [14] предложен подход на основе метода главных компонент. В начале для собранного массива мнений с использованием вероятностных критериев определяется ряд терминов для анализа, затем с использованием метода главных компонент выполняется переход к двумерному пространству признаков, которое имеет вид, представленный на рисунке 1.12.

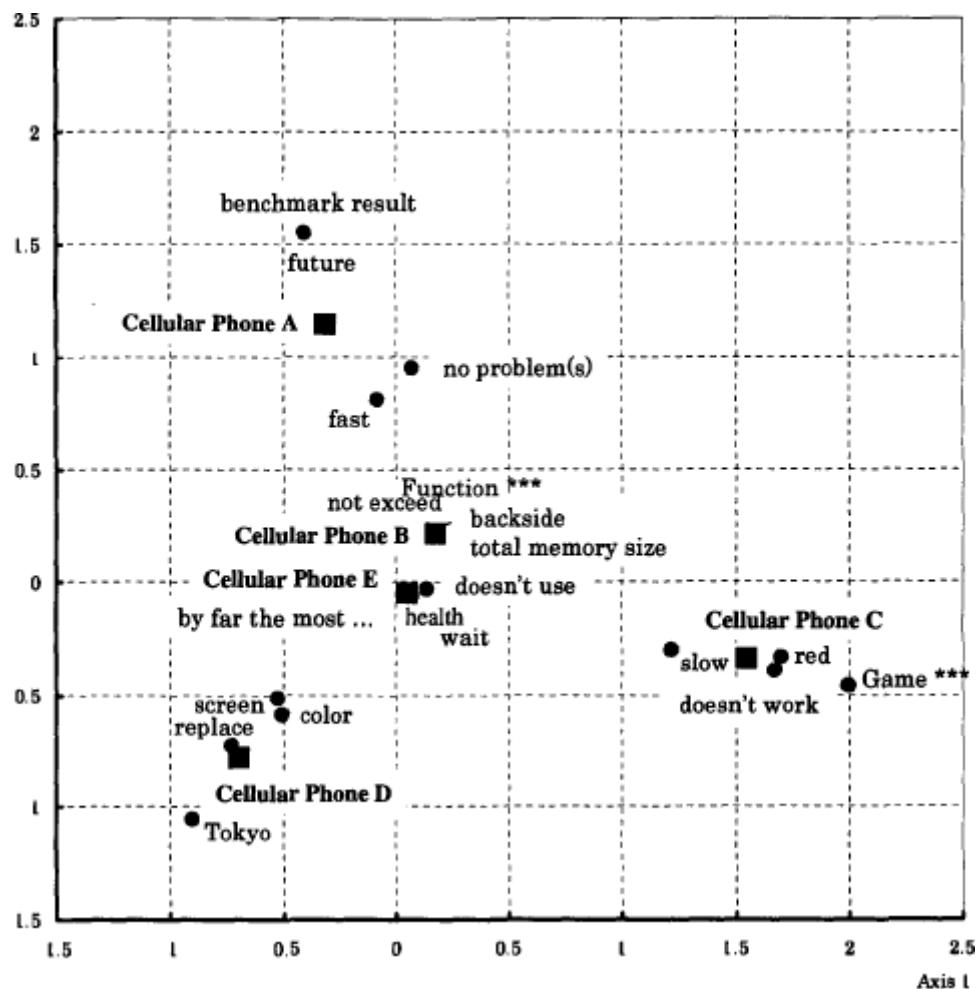


Рисунок 1.12 – Визуальная интерпретация мнений с использованием метода главных компонент

В полученном двумерном пространстве обсуждаемые объекты оказываются рядом с наиболее характеризующими их тонально окрашенными терминами.

Обобщая описанные в доступной литературе методы обработки мнений, можно выделить их следующие основные особенности:

- в основе методов обработки мнений лежат алгоритмы классификации текстов по тональности и выделения тонально окрашенной лексики;
- принятие решения в отношении тональности крупных блоков текста осуществляется на основе статистики тональности их отдельных элементов (слов, предложений и т. п.);
- визуализация массивов мнений осуществляется по результатам их статистической обработки с использованием геометрических пропорций и цветографической разметки;
- обобщение мнений выполняется в основном с использованием методов реферирования текстов, применяемых как ко всему массиву мнений, так и к его фрагментам, представляющим отдельные тематические и/или тональные рубрики;
- перспективные методы обобщения мнений могут быть основаны на формировании выборок, учитывающих наличие в массиве разных мнений;
- существующие исследования посвящены в основном анализу отзывов интернет-пользователей в отношении товаров и услуг, а также высказываний в средствах массовой информации в отношении различных событий и персон;
- не выявлено готовых моделей и методов обработки текстовых комментариев к законопроектам. Между тем такие модели и методы должны учитывать общие закономерности подходов к обработке мнений: выделение тематических, тональных и стилевых классов, применение обобщения мнений в классах с использованием реферирования, разработку перспективных методов обобщения с использованием выборок.

1.2.3 Выбор метода представления и обработки комментариев

Анализ процесса обработки комментариев к законопроектам экспертами и существующих методов обработки мнений показал, что в исследуемой предметной области существует противоречие, включающее теоретический и практический аспекты.

Теоретическое противоречие заключается в несоответствии между потребностью в моделях аналитической обработки комментариев в задаче обработки результатов обсуждения законопроектов и отсутствием качественной модели, пригодной для такой обработки. Практическое противоречие состоит в несоответствии между потребностью экспертов в аналитической обработке большого объема комментариев и ограниченными временными и человеческими ресурсами для обработки комментариев (рисунок 1.13).

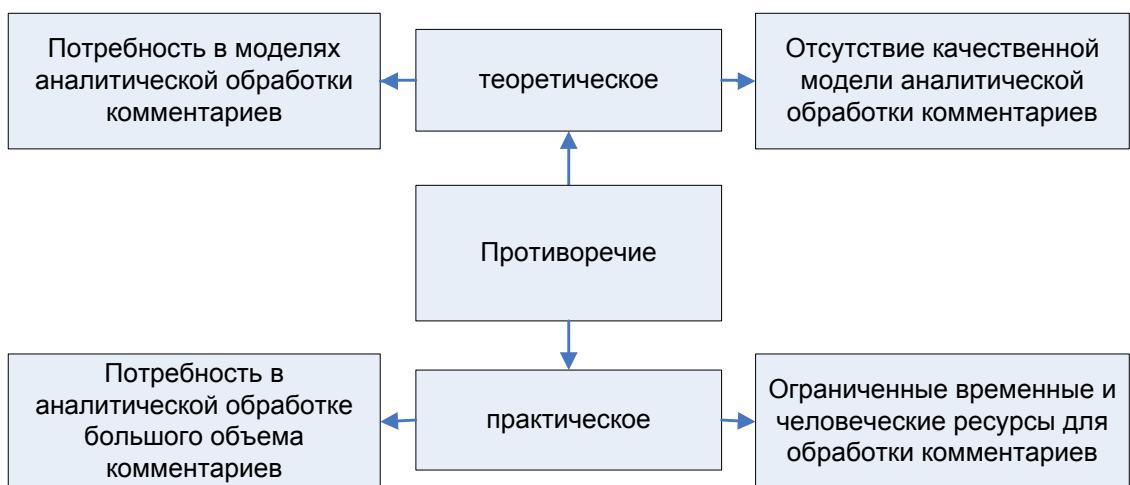


Рисунок 1.13 – Противоречие в исследуемой предметной области

Разрешение указанного противоречия целесообразно проводить путем моделирования действий эксперта по обработке комментариев с использованием моделей представления знаний [49–56].

Для выбора модели представления знаний, пригодной для обработки текстовых комментариев к законопроектам, проанализированы следующие типы моделей:

- семантические сети (СС) [57–60];
- логико-лингвистические модели (ЛЛМ) [61, 62];
- информационно-поисковые тезаурусы (ИПТ) [63, 64];
- фреймовые модели (ФМ) [65–67];
- онтологии (ОЛ) [68–72];
- модели оперативной аналитической обработки текстов (ОАОТ) [20–31].

Результат сравнения свойств данных моделей, наиболее полно отражающих особенности текстовых комментариев и процесса их обработки, приведен в таблице 1.3.

Таблица 1.3 – Выбор модели представления знаний, наиболее полно отражающей свойства текстовых комментариев к законопроектам и процесса их обработки

Свойства комментариев и процесса их обработки	Типы моделей					
	СС	ЛЛМ	ИПТ	ФМ	ОЛ	ОАОТ
Свойства комментариев						
Многомерность	0	0	1	1	1	1
Интерпретируемость	1	1	1	1	1	1
Типичность	1	0	0	0	0	1
Полезность	1	0	0	0	0	1
Свойства процесса обработки комментариев						
Многоуровневость	0	0	1	1	0	1
Масштабируемость	1	0	1	1	1	1
Динамичность	0	1	0	0	0	1
Визуальность	1	0	1	0	1	1
Интерактивность	1	0	0	1	1	1
Общая оценка, баллов	6	2	5	5	5	9

По совокупности характеристик, отражающих существенные свойства текстовых комментариев к законопроектам и процесса их обработки, предпочтительными являются модели оперативной аналитической обработки текстов. Проведем анализ существующих моделей данного типа.

1.2.4 Анализ моделей оперативной аналитической обработки текстов

В [21] рассматривается текстовый куб с операцией расчета информационно-поисковых мер для анализа многомерной текстовой базы данных.

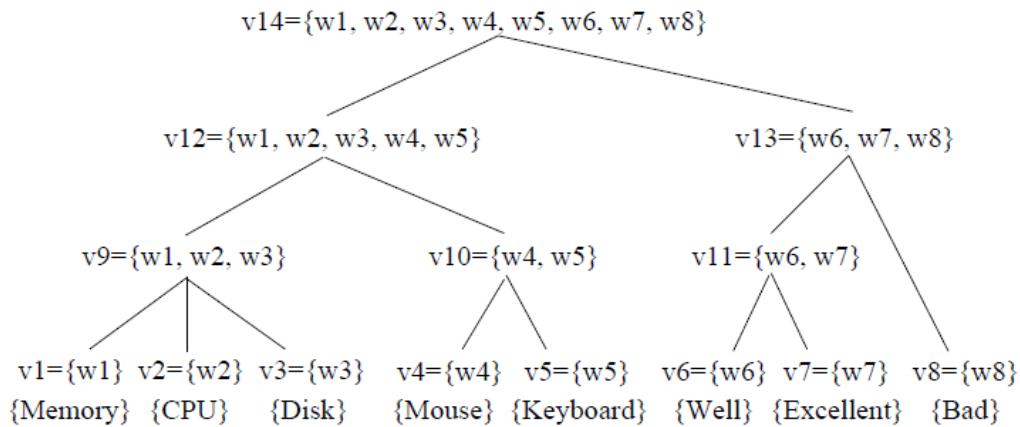
В 2008 г. С. Лин (C. Lin) и соавторы предложили модель куба данных, которая интегрирует возможности традиционного OLAP и техник информационного поиска, предназначенных для текста. В предложенной модели были объединены два типа иерархий: иерархия измерений и иерархия терминов. Экспериментальные результаты получены путем анализа данных пользовательских обзоров компании Dell.

Описание модели текстового куба включает следующие компоненты: многомерную модель данных, в которой поддерживаются два типа иерархий понятий (иерархия измерений и иерархия терминов), и меры, обеспечивающие проведение эффективного информационного поиска (частота термина и инвертированный индекс).

Поскольку иерархия измерений аналогична традиционным кубам данных, каждое измерение может состоять из множества атрибутов и организовано в виде дерева или ориентированного графа без петель. К ней применимы четыре OLAP-операции: всплытие, погружение, сечение, построение подкуба.

Иерархия терминов T строится на основе множества терминов W для задания семантических уровней и связей терминов. Каждый узел v в T называют обобщенным термином, представленным подмножеством терминов, т. е. $v \subseteq W$. В частности, каждый лист является синглетоном $\{w_i\}$, содержащим один термин $w_i \in W$, а корень T представляет собой множество всех терминов W . Множество дочерних узлов для v обозначают $\text{cld}(v)$, а множество родительских узлов – $\text{des}(v)$.

Пример иерархии терминов $T=\{v_1, \dots, v_{14}\}$, построенной на основе множества терминов $W=\{w_1, \dots, w_8\}$, приводится на рисунке 1.14. Так, обобщенный термин $v_9=\{\text{Memory}, \text{CPU}, \text{Disk}\}$ представляет "internal device".

Рисунок 1.14 – Пример иерархии терминов T

При построении куба в качестве мер применяются частота термина TF и инвертированный индекс IV . При этом если ячейка куба содержит агрегированные текстовые данные D и $W = \{w_1, w_2, \dots, w_m\}$ – множество всех терминов, то вектор частот терминов представляет m -мерный вектор:

$$TF(D) = \langle TF(w_1, D), TF(w_2, D), \dots, TF(w_m, D) \rangle, \quad (1.8)$$

а инвертированный индекс – также m -мерный вектор:

$$IV(D) = \langle IV(w_1, D), IV(w_2, D), \dots, IV(w_m, D) \rangle. \quad (1.9)$$

Пример применения текстового куба показан в таблице 1.4. В ней выбрано измерение "модель" и показано сечение, содержащее 10 наиболее частых терминов в полях таблицы "преимущества" и "недостатки".

Таблица 1.4 – Сравнение Dell XPS M1730 и Inspiron 1420N

<i>model</i>	XPS M1730	Inspiron 1420N
<i>pros</i>	game great keyboard work video nice design amaze display speed	nice love color light wireless cheap driver webcam quick speedy
<i>cons</i>	hard expensive battery problem keyboard carry game heavy screen need	battery problem gb hibernate ram virus button pad ubuntu g

Также известен тематический куб с операцией тематического моделирования для OLAP в многомерных текстовых базах данных.

В 2009 г. Д. Жаном (D. Zhang) и соавторами была предложена модель тематического куба, в которой возможности OLAP объединены с вероятностным тематическим моделированием и применены для поля текстовых данных в многомерной текстовой базе данных.

В [22] в качестве примера рассматривается система ASRS – авиационная система сообщений о безопасности, являющаяся самым большим в мире хранилищем информации об авиационных происшествиях. База данных содержит как структурированные, так и текстовые данные (таблица 1.5).

Таблица 1.5 – Пример текстовой базы данных в ASRS

ACN	Time	Airport	...	Light	Narrative
101285	199901	MSP	...	Daylight	Document 1
101286	199901	CKB	...	Night	Document 2
101291	199902	LAX	...	Dawn	Document 3

Поскольку факторы аномальных ситуаций отсутствуют в структурированной части данных хранилища и скрыты в текстовых полях, для поддержки деятельности аналитика требуется организация обработки данных как с использованием OLAP, так и с применением методов контент-анализа текстов. Несмотря на то, что существующие возможности OLAP позволяют проводить операции погружения и всплытия для измерений, основанных на структурированных атрибутах, они, тем не менее, не обеспечивают аналитику возможность погружения и всплытия для текстового измерения в соответствии с иерархией тем, представленной на рисунке 1.15.

По мнению авторов [22], для интеграции контент-анализа текста с OLAP необходимо обеспечить поддержку следующих функций: построения иерархии тематического измерения, формирования меры текстового содержания, эффективной реализации.

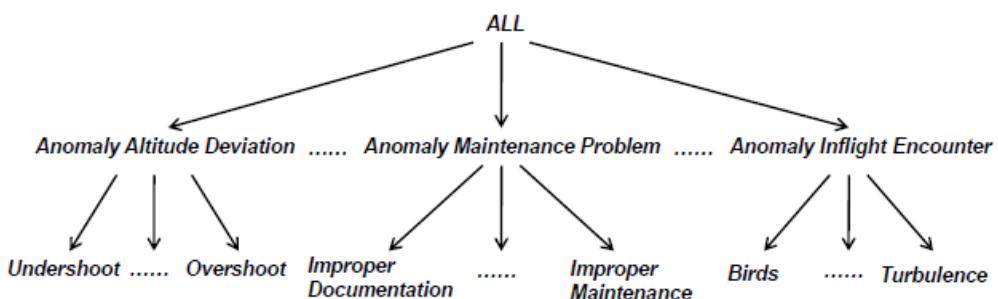


Рисунок 1.15 – Иерархическое дерево тем для аномальных событий

Пример тематического куба, построенного на данных ASRS, показан на рисунке 1.16. Измерения "время" и "место" представлены стандартными полями в текстовой базе данных ASRS, а измерение "тема" добавляется из иерархического дерева тем.

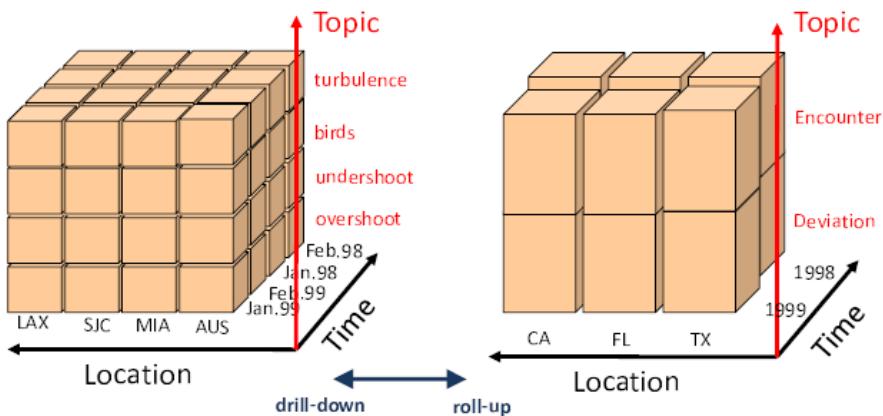


Рисунок 1.16 – Пример тематического куба

Пример двух ячеек тематического куба на базе ASRS, содержащих только меру "распределение слов", приведен в таблице 1.6. Когда эксперт проводит анализ с использованием тематического куба, он обладает дополнительными возможностями для понимания основных проблем при соответствующем аномальном событии.

Таблица 1.6 – Пример ячеек тематического куба

Time	Anomaly Event	Word Distribution
1999.01	equipment	engine 0.104, pressure 0.029, oil 0.023, checklist 0.022, hydraulic 0.020, ...
1999.01	ground encounters	tug 0.059, park 0.031, pushback 0.031, ramp 0.029, brake 0.027, taxi 0.026, tow 0.023, ...

Альтернативной моделью является информационный сетевой усовершенствованный текстовый куб.

В 2009 г. Й. Ю (Y. Yu) с группой ученых были рассмотрены подходы к анализу многомерных текстовых баз данных, позволяющие извлекать из них требуемые типы сущностей и взаимосвязи между ними.

В отличие от методов, требовавших привлечения экспертов для построения иерархий понятий, авторами предложены методы RankClus и NetClus для автоматического построения таких иерархий с использованием анализа информационных сетей [23]. При этом разработанный прототип объединяет применение OLAP и анализ информационных сетей в единую web-систему.

Для примера рассмотрены две многомерные текстовые базы данных: библиографическая база данных по компьютерным наукам конференции DBLP и новостная текстовая база данных из web.

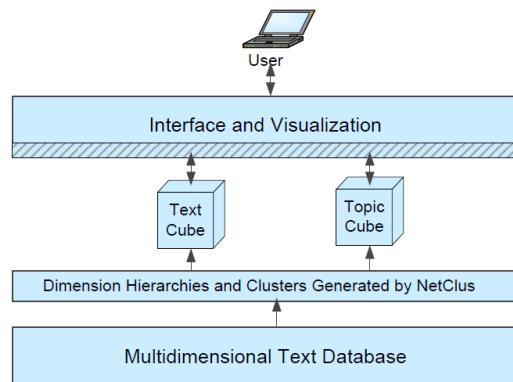


Рисунок 1.17 – Архитектура системы

Система имеет четырехуровневую архитектуру (рисунок 1.17). Нижний уровень представлен модулем NetClus, который анализирует информационные сети и генерирует кластеры, проводит ранжирование и создает иерархию понятий. Промежуточный слой включает текстовый и тематический кубы, которые обеспечивают оперативные информационно-поисковые меры и вероятностный скрытый семантический анализ. Верхний уровень предназначен для взаимодействия с пользователями.

При анализе базы DBLP модуль RankClus выделяет два типа сущностей: конференции и авторы. Используя экспертные правила, модуль RankClus выполняет одновременно кластеризацию и ранжирование. Модуль NetClus поддерживает многотипные (два и более типа) информационные сети для кластеризации, ранжирования и формирования иерархии понятий. Пример ранжирования всех типов сущностей для узла "Базы данных" приведен в таблице 1.7.

Таблица 1.7 – Ранжирование типов сущностей узла "Базы данных"

Conference	Rank Score	Author	Rank Score	Term	Rank Score
SIGMOD	0,315	Surajit Chaudhuri	0,0065	database	0,0529
VLDB	0,306	Michael Stonebraker	0,0062	system	0,0322
ICDE	0,194	C. Mohan	0,0053	query	0,0313
PODS	0,109	David J. DeWitt	0,0051	data	0,0251
EDBT	0,046	Jeffrey F. Naughton	0,0051	object	0,0138
CIKM	0,019	Michael J. Carey	0,0050	management	0,0113
...

Проведенная обработка делает возможной реализацию процедур поиска, OLAP-анализа и глубинного анализа в многомерной текстовой базе данных. Пример выполнения запроса показан в таблице 1.8.

Таблица 1.8 – К тем для запроса "graph" на выбранном пользователем измерении

Topic 1	Topic 2	Topic 3
mine	graphical	rank
large	graphics	query
index	language	efficient
search	interface	top
structure	design	answer
pattern	process	stream
fast	processor	integration
match	...	visual
...

После формирования иерархии понятий становится возможным применение OLAP-операций к многомерным текстовым базам данных. В отличие от традиционного куба данных, мерой данных кубов является текстовая мера в форме TF/IDF или тематической модели (распределения терминов).

Также известен документальный куб как средство многомерной визуализации и исследования больших множеств документов.

В 2003 г. Ж. Мозе (J. Mothe) с коллективом ученых представили новый пользовательский инструментарий, обеспечивающий визуализацию больших массивов документов для предоставления пользователям возможности формулирования запросов и доступа к документам [24]. Основой предлагаемого подхода является использование иерархий понятий для структуризации коллекции документов. Пользователям предоставляется интерфейс, основанный на принципах OLAP и операциях многомерного анализа, для исследования коллекции документов.

Предложенная модель документального куба основана на схеме "звезды" и состоит из нескольких таблиц измерений и таблицы фактов (рисунок 1.18).

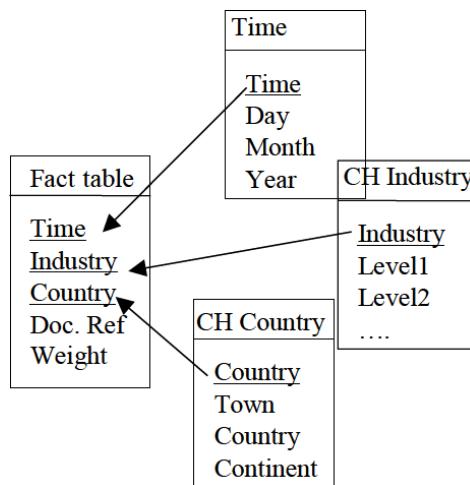


Рисунок 1.18 – Модель документального куба

Модель содержит одну таблицу для каждого измерения. Измерение организовано в соответствии с иерархией понятий, различные уровни представлены в одной таблице. Документальный куб обеспечивает двух- и трехмерные визуальные представления, характеризующие соответствие документов понятиям, выбранным пользователем (рисунок 1.19).

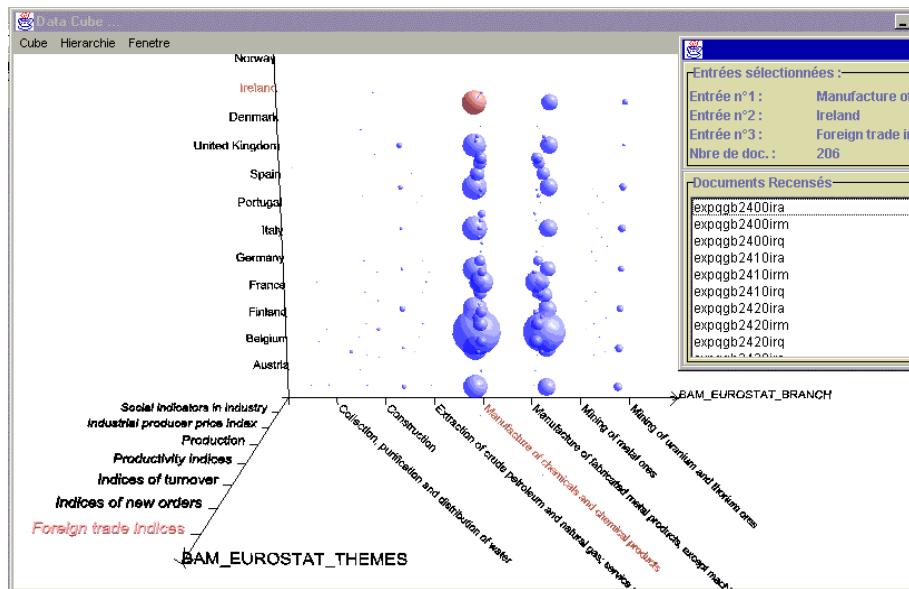


Рисунок 1.19 – Трехмерное представление документов

Оси куба соответствуют измерениям, тогда как сферы показывают число документов, попавших в соответствующую категорию. Размер сферы соответствует числу документов. При выполнении всплытия считается общее количество документов в дочерних узлах. Авторы рассматривают следующие операции с документальным кубом: всплытие и погружение, сечение (рисунок 1.20), доступ к документу.

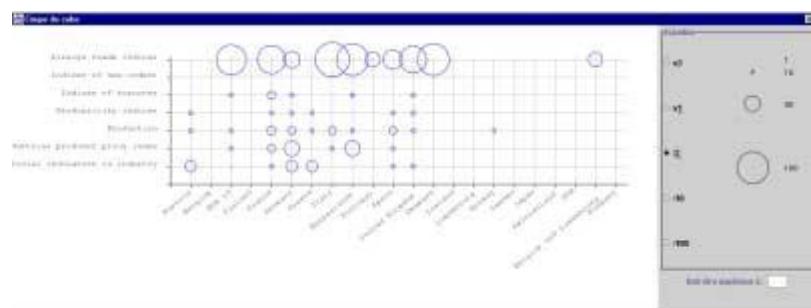


Рисунок 1.20 – Сечение документального куба

Прямой доступ может быть осуществлен путем выбора одной или нескольких сфер в кубе. Как результат, пользователю будет выдан ранжированный список ссылок на документы. Авторы применяют документальный куб для библиометрического анализа и мониторинга научных публикаций.

Также известны модели анализа больших коллекций электронных текстов с использованием OLAP.

В 2005 г. С. Кейт (S. Keith) с группой ученых предложили инструментарий для интерактивного, управляемого пользователем взаимодействия с "хранилищем слов" (рисунок 1.21), которое строится при помощи совокупности процедур извлечения, преобразования и загрузки.

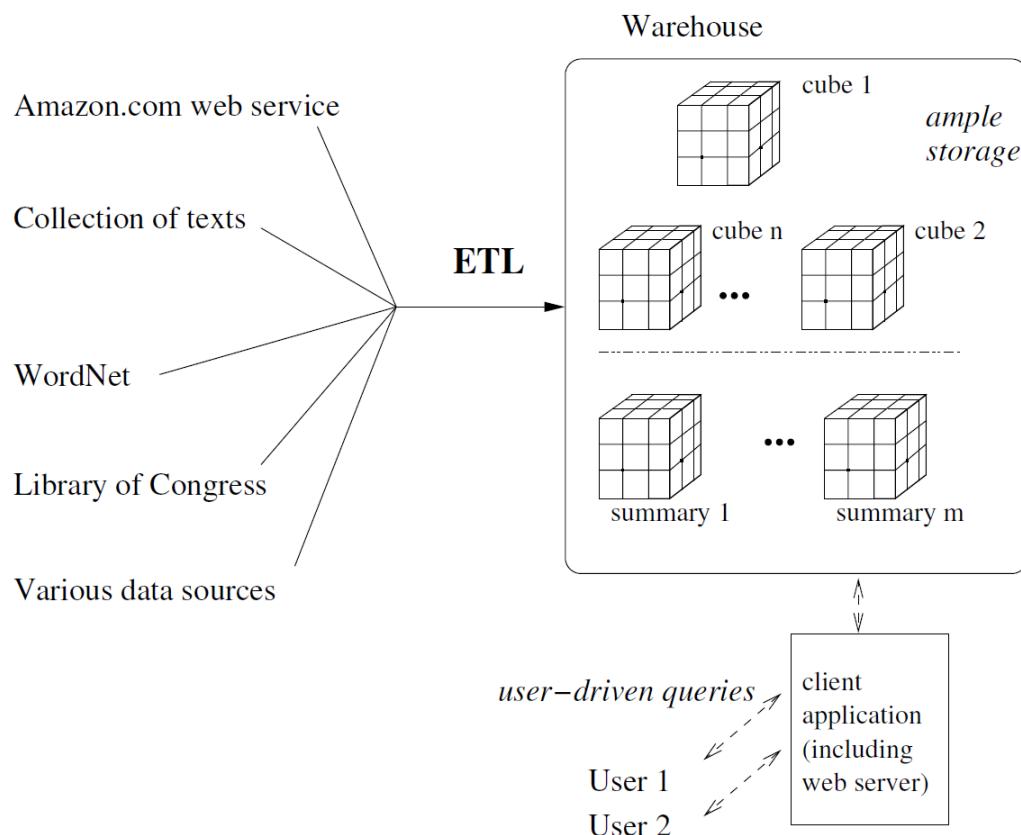


Рисунок 1.21 – Архитектура хранилища слов

Хранилище данных организовано с использованием традиционного куба данных, представляющего собой k -мерный массив, с несколькими определенными операциями, такими как: сечение, выделение подкуба, всплытие и погружение [25]. Пример 3-мерного куба данных с двумя измерениями "слово" и одним измерением "книга" приведен на рисунке 1.22.

Предлагается использовать хранилище слов для следующих приложений: определение авторства, лексический анализ, стилеметрический анализ, поиск похожих элементов.

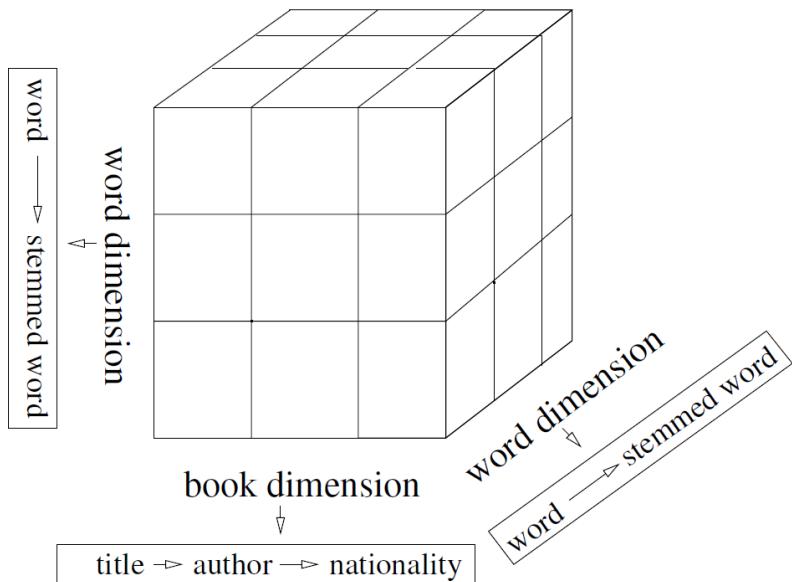


Рисунок 1.22 – Простой куб данных с тремя измерениями.

Для поддержки стилеметрии, поиска аналогий и анализа фраз хранилище слов использует две модели кубов: куб вида предложения и куб коротких фраз.

1. Куб вида предложения:

$$Book \times Word \times WordCount \times StopWordCount \rightarrow OccurrenceCount. \quad (1.10)$$

В данной формуле слева от знака \rightarrow через знак \times указаны измерения куба, а справа – содержимое ячейки куба. Каждый элемент формулы, содержащий "Count", имеет тип integer, а измерение "Word" представляет первое слово в предложении. При помощи данного куба исследуются характеристики предложений.

2. Куб коротких фраз:

$$Book \times Word \times Word \times Word \times Word \rightarrow OccurrenceCount. \quad (1.11)$$

Данный куб представляет все последовательности из четырех слов и может использоваться для исследования статистики общих или редких фраз по авторам или временным периодам.

Альтернативной моделью является текстовый куб для обнаружения документального подтверждения связи между сущностями.

В 2007 г. Х. Лоу (H. Lauw) с группой ученых предложили для обнаружения связей между сущностями и их документального подтверждения использовать подход на основе текстового куба [26].

В качестве примера авторы рассматривают задачу поиска связи между международной террористической организацией Аль-Каидой и Абу Сайяфом, группой сепаратистов на Филиппинах. Задача решается с использованием текстовых файлов о террористических актах, содержащихся в Базе знаний о терроризме (Terrorism Knowledge Base – TKB), www.tkb.org.

Пример текстового куба демонстрируется на рисунке 1.23. Он представлен как многомерная таблица с типами сущностей на измерениях. В данном случае на обеих осях представлены типы сущностей "организация". Затененные ячейки указывают на наличие хотя бы одного подтверждающего связь документа.

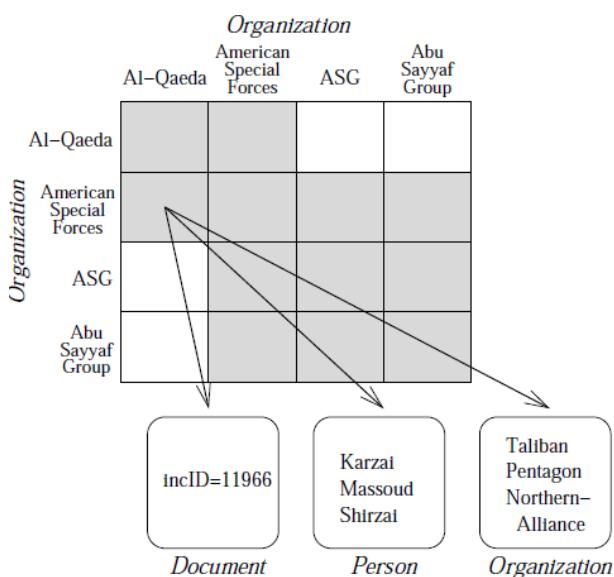


Рисунок 1.23 – Пример текстового куба

Примеры резюме ячеек включают список документов, подтверждающих связь, или список иных сущностей, имеющих отношение к данной связи. Пользователь взаимодействует с текстовым кубом посредством операций, проиллюстрированных на рисунке 1.24.

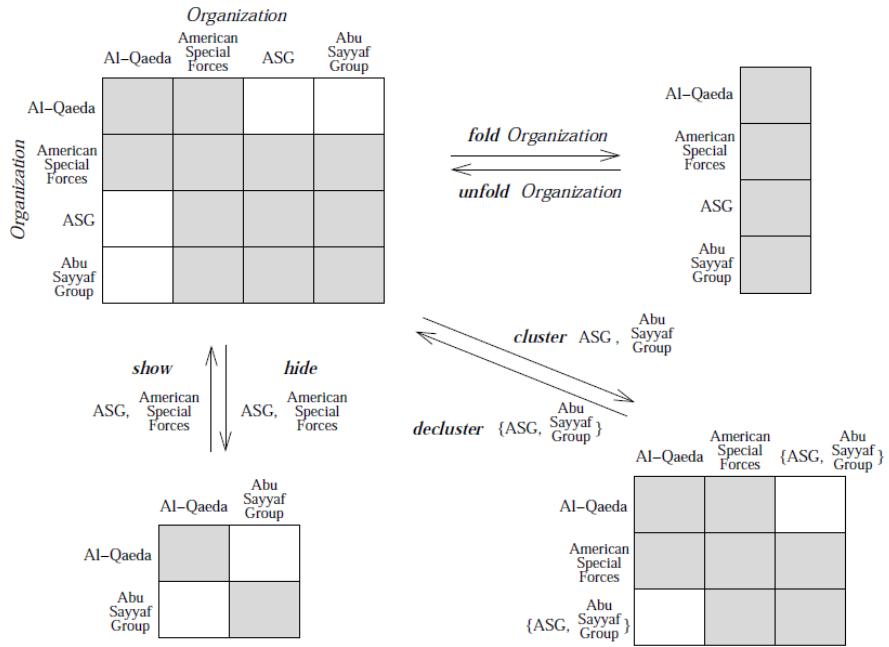


Рисунок 1.24 – Операции с текстовым кубом

Для вычисления силы связи между сущностями вводятся следующие обозначения: пусть A – множество документов, содержащих сущность a , а B – множество документов, содержащих сущность b . Тогда сила связи $link(a, b)$ может быть вычислена по формуле:

$$link(a, b) = \frac{|A \cap B|}{|A \cup B|}. \quad (1.12)$$

Распространяя данный подход на путь от a к d , состоящий из связей $\{(a, b), \dots, (c, d)\}$, обозначенный как $path(a, d)$, имеем:

$$path(a, d) = link(a, b) \times \dots \times link(c, d). \quad (1.13)$$

В расчетах используются пути длиной не более трех связей до целевой сущности. Также исключаются пути, проходящие через сущности, встречающиеся в пределах одного документа. После заполнения текстовый куб используется для интерпретации связей.

Также известен кубический индекс как модель индекса текста для поиска и глубинного анализа.

В 2010 г. ученые из Индии Б. Жанет и А. Редди (B. Janet, A. Reddy) предложили индекс, использующий модель текстового куба для хранения текстовой информации по аналогии с кубом данных в Data Mining [27]. Модель включает три типа кубов: прямой индекс, индекс следующего слова и инвертированный индекс в едином трехмерном кубическом индексе. В качестве измерений рассматриваются слово, следующее слово и документ. Мерой в кубе является частота встречаемости пары "слово-следующее слово".

Кубический индекс позволяет моделировать и просматривать текст во множестве измерений. В качестве фактов выступают частоты терминов или другие взвешивающие термины меры.

Для примера рассмотрен текст "Magnetic Lines of Force acting on earth". Слова в данной фразе обозначены как w_1, w_2, w_3, w_4, w_5 . Слово n_1 является следующим для w_1 (Lines для Magnetic), n_2 – для w_2 и т. д. Частота встречаемости фразы равна единице. Куб для данного примера показан на рисунке 1.25.

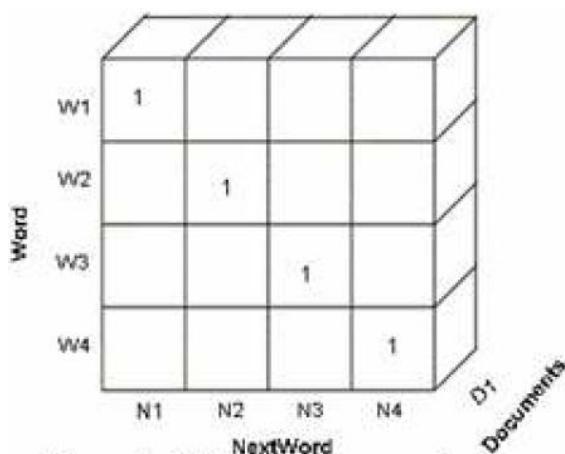


Рисунок 1.25 – Конструкция кубического индекса

Куб представляет связи слов в пространстве "слово-следующее слово", а в пространстве "слово-документ" – связи слов и документов.

Кубический индекс также может использоваться для разработки ассоциативных правил между словами для усиления механизма запросов за счет применения априорного алгоритма.

В качестве иерархии понятий для измерения слов рассматриваются фраза, предложение, параграф или блок рассматриваемого текста. В кубическом индексе применимы стандартные OLAP-операции: всплытие, погружение, сечение, определение подкуба, вращение.

Авторы описывают структуру скрытого семантического индекса, получаемого в пространстве "слово-документ". Данный индекс может использоваться для поиска ассоциаций слов и документов.

При выделении меры для информационного поиска и глубинного анализа используются следующие обозначения: пусть N – общее число представленных документов, F_{11} – количество появлений слова w_1 в документе D_1 , n_1 – число документов, в которых встречается слово w_1 , тогда:

- частота термина = F_{12} ;
- инвертированная частота документа = $\log(N/n_1)$;
- вес термина = $F_{12} \cdot \log(N/n_1)$.

В экспериментах авторы рассматривают в качестве меры частоту термина. Для реализации кубического индекса применено программное обеспечение Terrier 2.1. Построены следующие виды индексов: лексический, индекс документа, прямой индекс следующего слова, кубический индекс.

В сводном виде характеристики перечисленных моделей представлены в таблице 1.9.

Для данных моделей характерны следующие общие признаки:

- модели представляют собой расширение классического куба данных для обработки текстов;
- извлечение семантики отнесено к этапу работы с построенным гиперкубом;
- категории, как правило, представлены иерархиями тем или понятий, выделенных из текста;

Таблица 1.9 – Характеристики существующих моделей оперативной аналитической обработки текстов

Авторы	Страна, год	Предмет анализа	Измерения	Мера	Операции анализа	Что введено в OLAP для текстов
Duo Zhang, ChengXiang Zhai, Jiawei Han, Ashok Srivastava, Nikunj Oza	США 2009	ASRS (Система авиационных отчетов о безопасности)	Тема из иерархического дерева тем	Распределение слов (unigram language model)	Drill-down, roll-up	Иерархия категорий-тем, операция погружения с использованием списка ключевых слов и их частот
Hadly W. Lauw, Ee-Peng Lim, Hwee-Hwa Pang	Сингапур 2007	Файлы инцидентов из ТКВ (Terrorism Knowledge Base)	Типы сущностей: персоны, организации	Ассоциация между двумя сущностями, итог ячейки, список документов (ранжированный или кластеризованный), список связанных сущностей	Fold/un-fold, show/hide, cluster/decluster, summarization	Операция погружения с использованием ключевых понятий
Cindy Xide Lin, Bolin Ding, Jiawei Han, Feida Zhu, Bo Zhao	США 2008	База отзывов покупателей компании Dell	Иерархия измерений, иерархия терминов	Частота термина (TF), инвертированный индекс (IV)	Roll-up, drill-down, slice, dice, pull up, push down	Иерархия категорий-терминов, операция погружения с использованием ключевых слов
Yintao Yu, Cindy X. Lin, Yizhou Sun, Chen Chen, Jiawei Han, Binbin Liao, Tianyi Wu, ChengXiang Zhai, Duo Zhang, Bo Zhao	США 2009	Библиографическая база по компьютерным наукам конференций DBLP, новостная текстовая база, извлеченная из web	База DBLP: конференции, авторы, статьи, термины. Новостная база данных: люди, место, время	Текстовый куб: TF, IDF. Тематический куб: узел иерархического дерева приоритетов	Roll-up, drill-down, slice, dice	Операция погружения с использованием списка ключевых слов или понятий заданного типа и их частот
B. Janet, A.V.Reddy	Индия 2010	Коллекция документов	Первое слово, следующее слово, документ, фраза/предложение/параметр/блок текста	Частота встречаемости фразы (пары слово-следующее слово)	Roll up, drill down, slice, dice, pivot (rotate)	Измерения представлены списком слов текста
Steven Keith, Owen Kaser, Daniel Lemire	Канада 2005	Плоский текст и XML-документы проекта Gutenberg	Книга/глава, год публикации/литературная эра/декада/век, слово/конечный суффикс/часть речи	Частота слова, частота пунктуации, длина предложений	Slice, dice, roll-up, drill-down	Измерения представлены списком объектов (слов, знаков препинания и т. д.) из текстов
Josiane Mothe, Claude Chrisment, Bernard Dousset, Joel Alaux	Франция 2003	Коллекция документов, содержание документа	Иерархия понятий для каждого измерения. Мониторинг науки: авторы, организации, дата документа, страницы, продукты. Экономика: индикаторы, отрасль, регион	Ссылка на документ (идентификатор, URL), вес	Drill-down, roll-up, slice	Иерархия категорий-понятий, выделенных из текстов, ячейки-сферы, операция погружения с использованием ранжированного списка документов

- в большинстве задач метод OLAP применяется к гиперкубовой модели без задания определенной траектории перемещения, т. е. в классическом виде, в то время как в отдельных задачах для поиска закономерностей задается определенный порядок применения операций визуального анализа;
- меры рассчитываются в сильных шкалах на основе проведения контент-анализа;
- применяются в основном стандартные операции анализа, а вновь вводимые используют те или иные принципы агрегации данных.

Достоинством рассмотренных моделей является возможность описания массивов текстов, допускающих многомерное представление.

В то же время перечисленные модели обладают следующими недостатками:

- модели не универсальны и в каждой из практических задач имеют собственный набор атрибутов;
- отсутствуют операции погружения в текстовые данные с обоснованной эффективностью.

Указанные факты свидетельствуют о недостаточной разработанности моделей оперативной аналитической обработки текстов применительно к мнениям, а в особенности – к текстовым комментариям, полученным в ходе общественного обсуждения законопроектов. Разрабатываемая модель должна учитывать следующие основные особенности текстовых комментариев к законопроектам, процедуры их обработки, а также тенденции развития методов обработки мнений:

- возможность оценки интенсивности мнений в классах, сформированных значимыми атрибутами;
- возможность обобщения мнений на основе рефератов и выборок;
- возможность ознакомления с типичными и полезными комментариями в соответствии с выделенным на обработку бюджетом времени.

1.2.5 Концепция системы обработки комментариев

Общий замысел информационной поддержки принятия решений по использованию комментариев для улучшения законопроектов состоит в снижении энтропии анализируемого множества текстов за счет трех основных процессов:

1. Структуризация массива комментариев на основе модели гиперкуба и предоставление инструментов навигации по измерениям гиперкуба с целью визуального анализа имеющихся зависимостей (использование технологии OLAP для оперативной аналитической обработки массива комментариев).

2. Автоматическое формирование выборки типичных комментариев для облегчения понимания экспертом сущности предложений, сгруппированных в кластеры подобных комментариев.

3. Автоматическая отсечка наименее полезных комментариев путем формирования методом направленного отбора выборки полезных комментариев для снижения объема анализируемых текстов.

В основе автоматического формирования выборки типичных комментариев лежит понятие конструктивной типологии.

Конструктивная типология [73] – это эвристический способ обработки и упорядочения информации, основанный на выборе или формировании типичного представителя – представителя типа какого-либо явления, а также распространении выводов, полученных при его анализе, на всю совокупность элементов данного типа.

Ценность типичного представителя как одного из компонентов знания определяется не точностью его соответствия реальности, а его способностью объяснить эту реальность.

В случае анализа комментариев к законопроектам полезность использования конструктивной типологии состоит в резком снижении объема анализируемых данных, необходимых для понимания экспертом основной проблематики.

В целом конструктивная типология есть средство сведения различий и разнообразия частных случаев к единому общему представителю типа, который не описывает конкретную структуру или действие, а подчеркивает ключевой атрибут или группу атрибутов, важных для понимания существа предложений или проблемы.

Характеристикой типичного представителя для объектов числовой природы, как правило, выступают средние различного типа, или геометрический центр класса в пространстве признаков, измеренных в сильных шкалах [74].

В качестве типичного представителя для кластера комментариев, выделенных методом "семантической" фильтрации [75] в задаче анализа комментариев к законопроектам, следует использовать или конкретный комментарий, наиболее близкий к центру кластера, или конструированный тип – реферат всех комментариев из одного кластера.

Элементами этого типичного представителя являются ключевые фразы из разных комментариев, объединенных в один кластер на основе "семантической" фильтрации.

Как логический прием, типичный представитель не претендует на эмпирическую достоверность, сохраняющую неповторимые аспекты каждого предложения; как концептуальный прием, для задачи обработки комментариев он представляет собой попытку перехода от описания конкретных предложений в комментариях к обобщенным представлениям о сущности замечаний по статьям законопроекта с целью лучшего понимания существа вопроса.

Типичный представитель сам по себе имеет характер теоретической модели, выступающей как объяснительная схема, обобщенная модель сущности предложений по законопроекту в каждом кластере предложений.

В основе автоматической отсечки наименее полезных комментариев путем формирования выборки полезных комментариев лежит их направленный отбор с использованием количественной оценки полезности по признакам атрибутов конструктивных предложений и предпочтений интернет-пользователей.

Методы, основанные на формировании выборок, широко применяются в различных исследованиях [76–78] в целях снижения объема материала, подлежащего обработке, при сохранении статистических свойств генеральной совокупности исследуемых объектов либо оптимизации характеристик группы объектов, включаемых в выборку.

Цель формирования выборки полезных комментариев состоит в представлении эксперту, осуществляющему выбор конструктивных предложений, возможности последовательного перехода от общих представлений о комментариях как совокупности неструктурированной информации к анализу ограниченного объема структурированных предложений с автоматической отсечкой наименее полезных комментариев.

Для достижения цели генеральная совокупность всех комментариев разбивается на когорты [79] – совокупности комментариев, включающих относительно однородные комментарии, объединенные по признакам: статья законопроекта, функциональный стиль текста, оценочная тональность текста, эмоциональная тональность текста.

По мощности когорт (интенсивности мнений) можно оценить заинтересованность в конкретной статье и тем самым проранжировать все статьи с точки зрения приоритетности анализа. Возможно проведение аналогичной ранжировки по всем типам когорт.

Внутри когорт комментарии ранжируются по обобщенному признаку, характеризующему наличие атрибутов конструктивных предложений и предпочтений интернет-пользователей. Принадлежность комментария к низким рангам в когорте может служить основанием для исключения его из рас-

смотрения как малоконструктивного при наличии временных ограничений на обработку комментариев.

Реализация общего замысла информационной поддержки принятия решений по использованию комментариев для улучшения законопроектов осуществляется на основе технологий системы поддержки принятия решений по использованию комментариев интернет-пользователей к правовым актам, представленных на рисунке 1.26 в форме функциональной схемы.

Базовые технологии компьютерной лингвистики, включающие морфологический и синтаксический анализ, решают задачи выделения морфологических токенов и минимальных смысловых фрагментов [50, 52, 64], необходимых для формирования выборок типичных и полезных комментариев.

Выборка типичных комментариев формируется на основе методов "семантической" фильтрации и реферирования, которые обеспечивают кластеризацию комментариев и формирование типичного представителя в каждом кластере в форме сводного реферата комментариев кластера с последующим выбором одного из комментариев данного кластера.

Выборка полезных комментариев формируется на основе классификации комментариев как прикладной технологии АОТ, оценки полезности отдельных комментариев на основе признаков, характеризующих наличие атрибутов конструктивных предложений и предпочтений интернет-пользователей, методов оптимизации, учитывающих разбиение по классам и убывающую полезность комментариев в классах.

Технология атрибутивного анализа обеспечивает выделение в комментариях атрибутов, характеризующих конструктивные предложения пользователей. Основными выделяемыми атрибутами являются: "дополнить", "изменить", "удалить" с их синонимами.

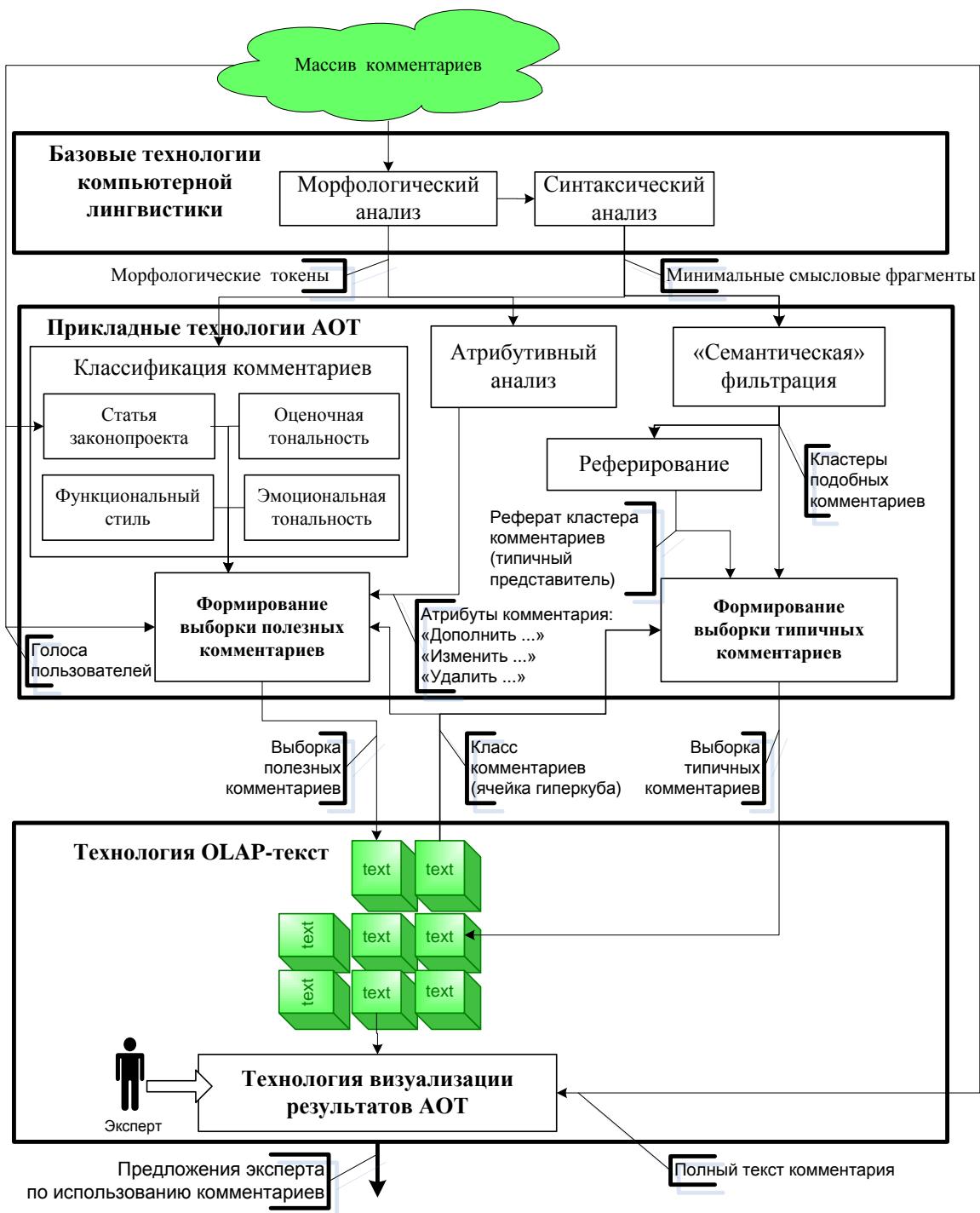


Рисунок 1.26 – Функциональная схема системы поддержки принятия решений по использованию комментариев к правовым актам

Анализ интенсивности мнений в классах производится с использованием OLAP-технологии (рисунок 1.27), дополненной операциями погружения в массив комментариев на основе формирования выборок типичных и полез-

ных комментариев в выбираемых аналитиком классах (разновидность технологии OLAP-текст).

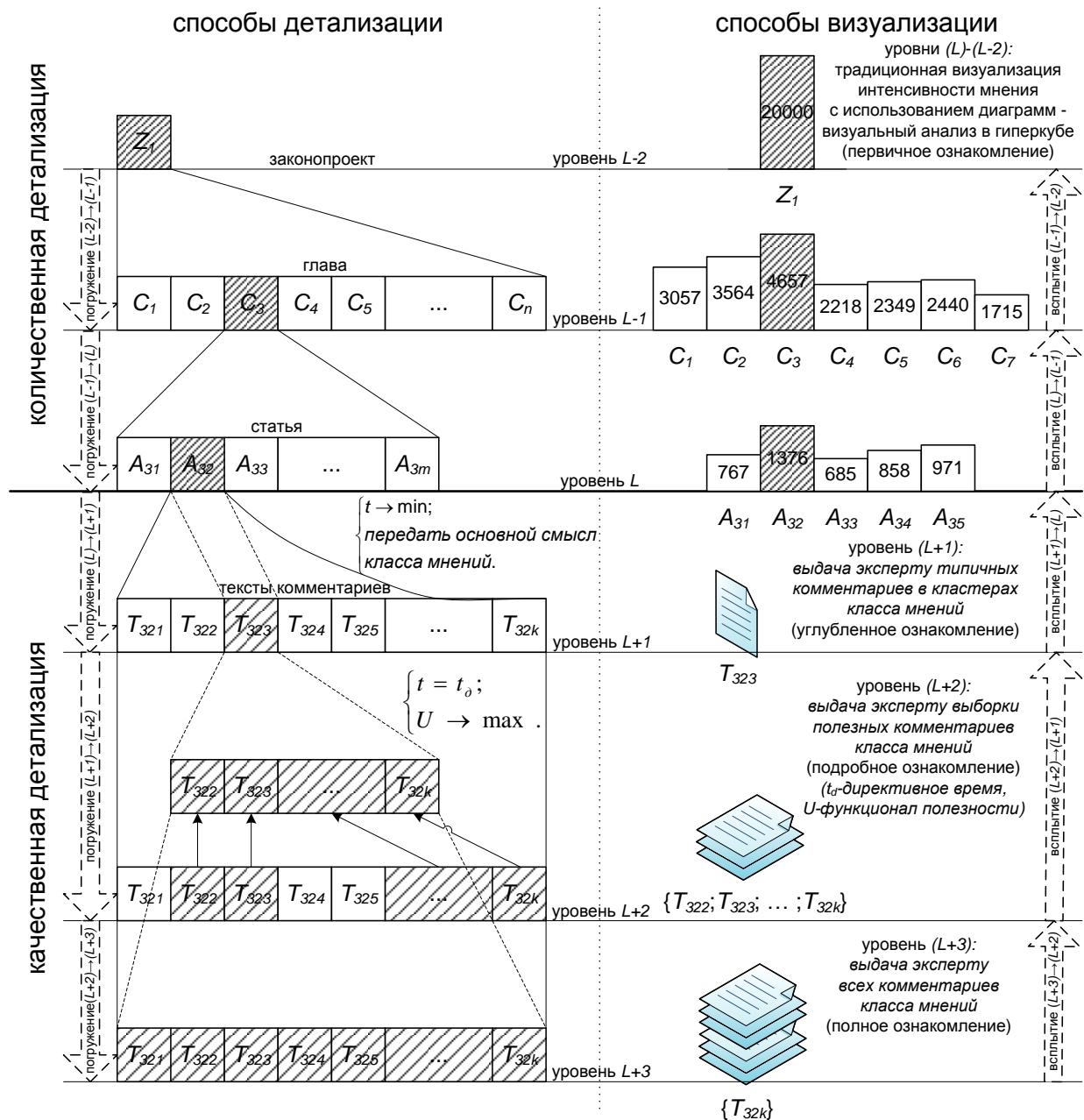


Рисунок 1.27 – Способы детализации и визуализации при погружении в массив комментариев в ходе оперативной аналитической обработки мнений

Такая система обработки комментариев позволит эксперту понять сущность предложений интернет-пользователей по корректировке обсуждаемого законопроекта и сформировать предложения по использованию комментариев.

1.3 Постановка задачи исследования

Для формализации задачи исследования используется теоретико-множественное представление.

Дано:

$SENT = \bigcup_{m=1}^n sent_m$ – множество высказываний.

$COM = \bigcup_{j=1}^k com_j$ – множество комментариев.

$L = \{l_j\}$ – множество весов комментариев, определяемых количеством голосов респондентов ("лайков"), отданных за комментарий, $j \in \overline{1, k}$.

$A = \bigcup_{s=1}^z a_s$ – кортеж атрибутов комментариев, где:

z – мощность кортежа, $z = 7$;

a_1 – глава законопроекта, к которой относится комментарий, $a_1 \in \overline{1, a_{1_{\max}}}$;

a_2 – статья законопроекта, к которой относится комментарий, $a_2 \in \overline{1, a_{2_{\max}}}$;

a_3 – идентификатор комментария, $a_3 \in \overline{1, a_{3_{\max}}}$;

a_4 – функциональный стиль комментария, $a_4 \in \{\text{научный, официаль-но-деловой, газетно-публицистический, художественно-беллетристический, разговорный}\}$;

a_5 – оценочная тональность комментария, $a_5 \in \{\text{мажорная, минорная}\}$;

a_6 – эмоциональная тональность комментария, $a_6 \in \{\text{высокая, средняя, низкая}\}$;

a_7 – количество комментариев, $a_7 > 0$.

$\Omega = \bigcup_{v=1}^w \omega_v$ – множество классов комментариев, формируемых с использованием декартова произведения подмножества атрибутов из множества A , причем число классов ω_v определяет число ячеек гиперкуба.

$\Psi = \bigcup_{j=1}^k \psi_j$, где $\psi_j : SENT \rightarrow COM$ – множество функций синтеза комментариев респондентами.

Действие функций ψ_j целесообразно рассматривать с учетом разбиения множества комментариев на классы ω_v , при этом необходимы следующие уточнения:

$$1. SENT_v = \bigcup_{m_v=1}^{n_v} sent_{v,m_v}, \bigcup_{v=1}^w SENT_v = SENT.$$

$$2. COM_v = \bigcup_{j_v=1}^{k_v} com_{v,j_v}, \bigcup_{v=1}^w COM_v = COM.$$

$$3. \Psi_v = \bigcup_{j_v=1}^{k_v} \psi_{v,j_v}, \bigcup_{v=1}^w \Psi_v = \Psi.$$

$$4. L_v = \{l_{v,j_v}\}.$$

Тогда процесс синтеза комментариев в ходе обсуждения законопроекта представляется как последовательное добавление вершин $sent_{v,m_v}$, com_{v,j_v} и ребер $(sent_{v,m_v}, com_{v,j_v})$, где $v \in [1, w]$, $m_v \in [1, n_v]$, $j_v \in [1, k_v]$ в двудольный граф $G = (\{SENT_v\}, \{COM_v\}; \{(sent_{v,m_v}, com_{v,j_v})\})$ при добавлении в множество Ψ очередной функции ψ_{v,j_v} .

При этом для добавленных вершин com_{v,j_v} производится взвешивание, в ходе которого определяются веса l_{v,j_v} (рисунок 1.28).

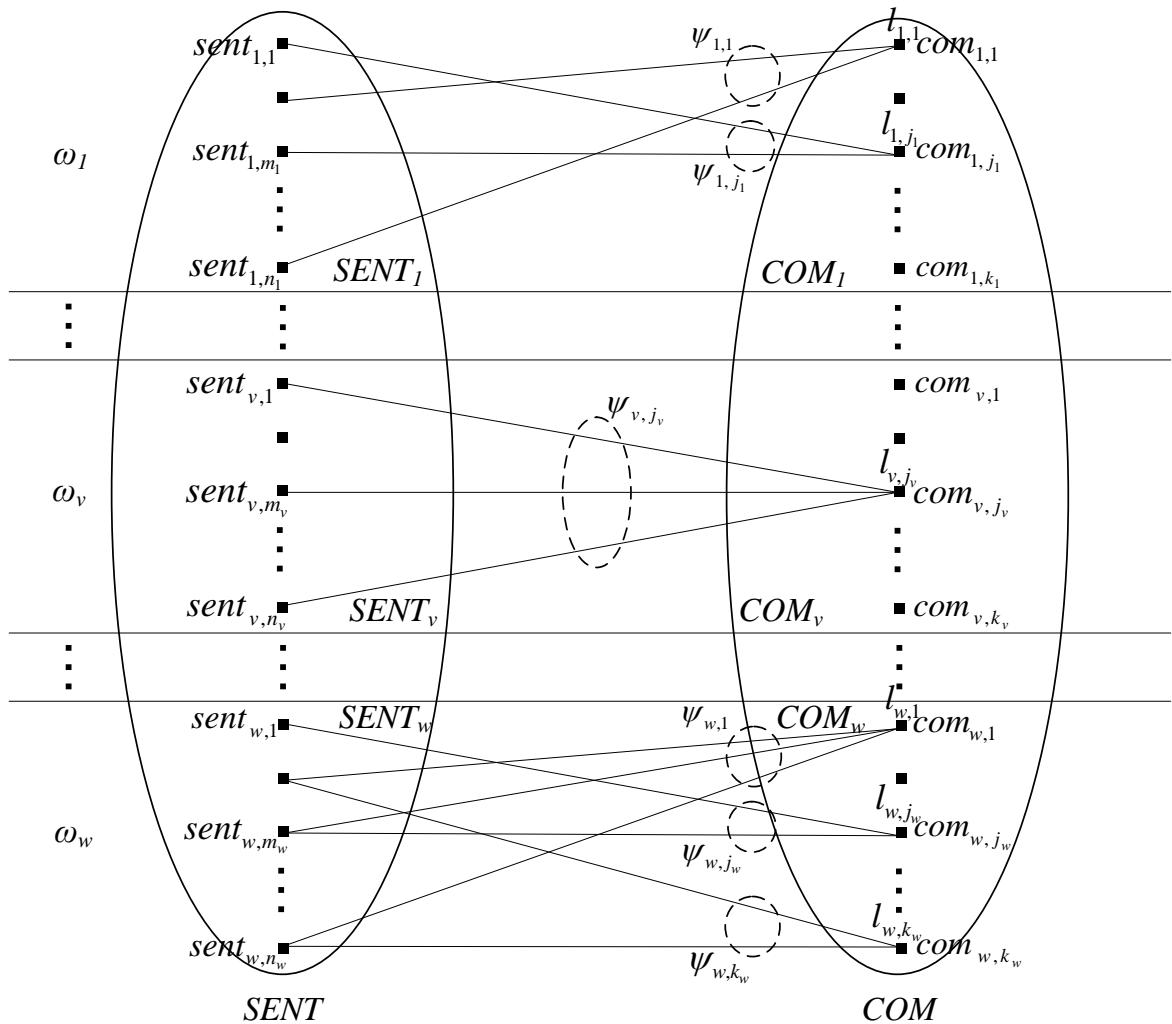


Рисунок 1.28 – Граф процесса синтеза комментариев
в ходе обсуждения законопроекта

При этом $\forall v$ области определения функций ψ_{v,i_v} и ψ_{v,j_v} , $i_v \neq j_v$, взаимодействуют одним из следующих способов:

$$1. \exists i_v, j_v \in [1, k_v], i_v \neq j_v: D_{\psi_{v,i_v}} \cap D_{\psi_{v,j_v}} = \emptyset.$$

$$2. \exists i_v, j_v \in [1, k_v], i_v \neq j_v: D_{\psi_{v,i_v}} \cap D_{\psi_{v,j_v}} \neq \emptyset.$$

Результат синтеза комментариев в ходе обсуждения законопроекта представляет собой бинарное отношение $\Psi \subset SENT \times COM$, причем $\forall v: \Psi_v \subset SENT_v \times COM_v$ (рисунок 1.29).

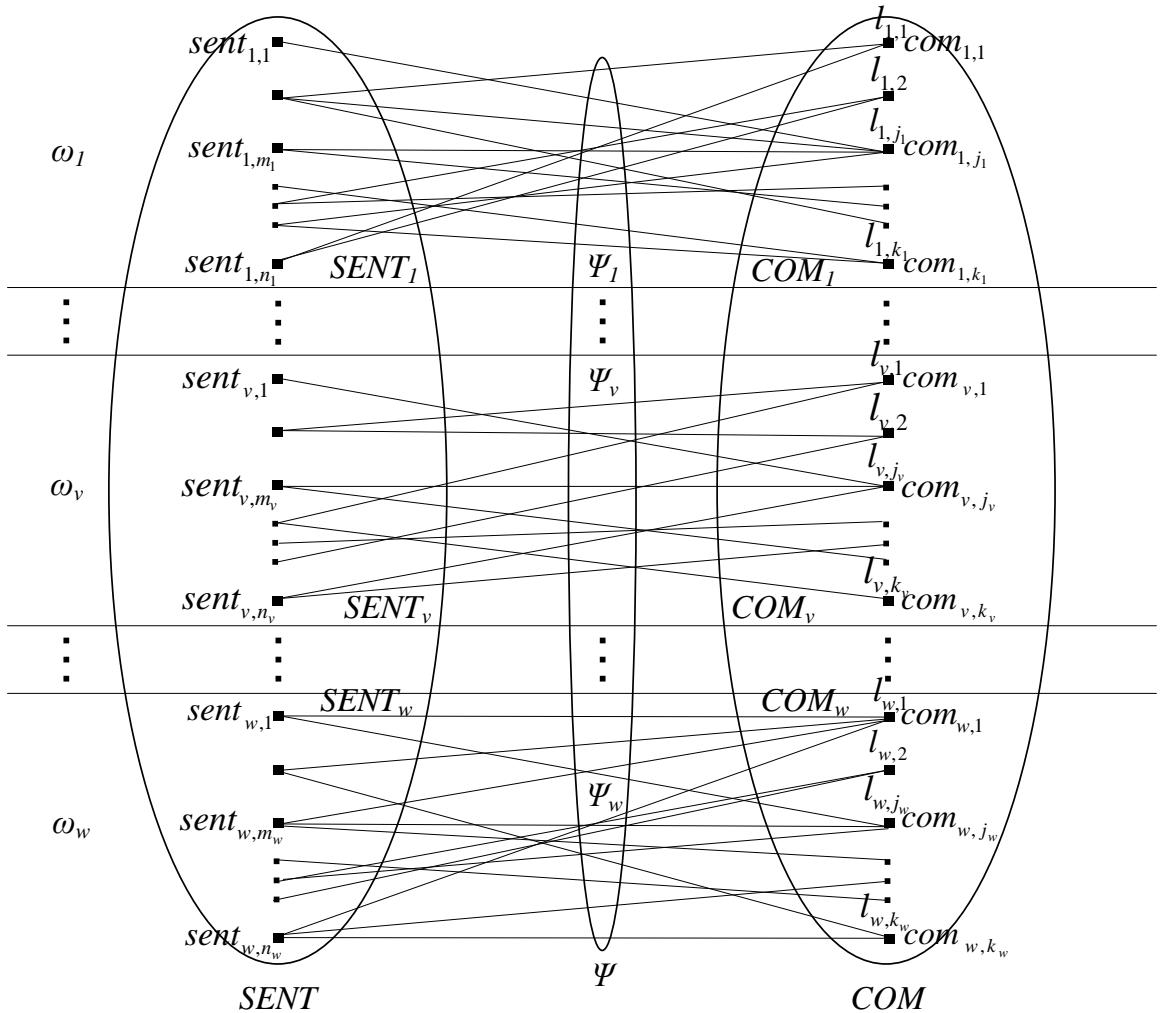


Рисунок 1.29 – Граф результата синтеза комментариев
в ходе обсуждения законопроекта

Ограничения и допущения:

1. $\forall v, q \in [1, w], v \neq q : SENT_v \cap SENT_q = \emptyset$.
2. $\forall v, q \in [1, w], v \neq q : COM_v \cap COM_q = \emptyset$.

Требуется:

для графа $G = (\{SENT_v\}, \{COM_v\}; \{sent_{v,m_v}, com_{v,j_v}\})$ предложить модель обработки элементов множеств $\{COM_v\}$ при заданном директивном времени обработки Td на основе определения подмножеств $\{com_{v,j_v}\} \subseteq \{COM_v\}$, обеспечивающую выполнение требования:

$$\begin{cases} \frac{\overline{\Delta R_{\{com_{v,j_v}\}}}}{\Delta t} \geq \frac{\overline{\Delta R_{\{COM_v\}}}}{\Delta t}, \\ 0 \leq t \leq Td, \end{cases} \quad (1.14)$$

где $R_{\{com_{v,j_v}\}} = \frac{|SENT_{\{com_{v,j_v}\}}|}{|SENT_v|}$ – полнота присутствия высказываний в отобранных комментариях множества $\{com_{v,j_v}\}$ в сравнении с комментариями

множества COM_v , $R_{\{COM_v\}} = \frac{|SENT_{\{COM_v\}}|}{|SENT_v|}$ – полнота присутствия высказываний в последовательно взятых за бюджет времени Td комментариях исходного множества COM_v , t – время.

При выборе для решения данной задачи оперативной аналитической обработки текстов требуют решения следующие подзадачи:

1. Сформировать на основе множества значимых атрибутов $A = \bigcup_{s=1}^z a_s$ гиперкубовую модель $hc: a_1 \times \dots \times a_s \rightarrow a_r$, $s, r \in [1, z]$, $s \neq r$ и описать инъективные функции детализации $f_{TCS}: COM \rightarrow TCS$ и $f_{UCS}: COM \rightarrow UCS$ для обеспечения операции качественного погружения в процессе оперативной аналитической обработки массива текстовых комментариев.

2. Пусть $y_{v,m_v} = d(sent_{v,m_v})$ – количество комментариев, содержащих высказывание $sent_{v,m_v}$, определяемое степенью соответствующей вершины графа G , y_{v,m_v} характеризует типичность высказывания $sent_{v,m_v}$.

Необходимо найти такое множество комментариев $TCS = \bigcup_{v=1}^w TCS_v$, $TCS_v = \{com_{v,j_v}\}$, $TCS \subseteq COM$, представляющее собой выборку типичных комментариев, для которого выполняется система требований:

$$\begin{cases} |TCS| \rightarrow \min; \\ \sum_{v=1}^w \sum_{j_v=1}^{g_v} y_{v,j_v} \rightarrow \max; \\ Td = \text{const}, \end{cases} \quad (1.15)$$

где $g_v = |TCS_v|$ – мощность множества TCS_v (рисунок 1.30).

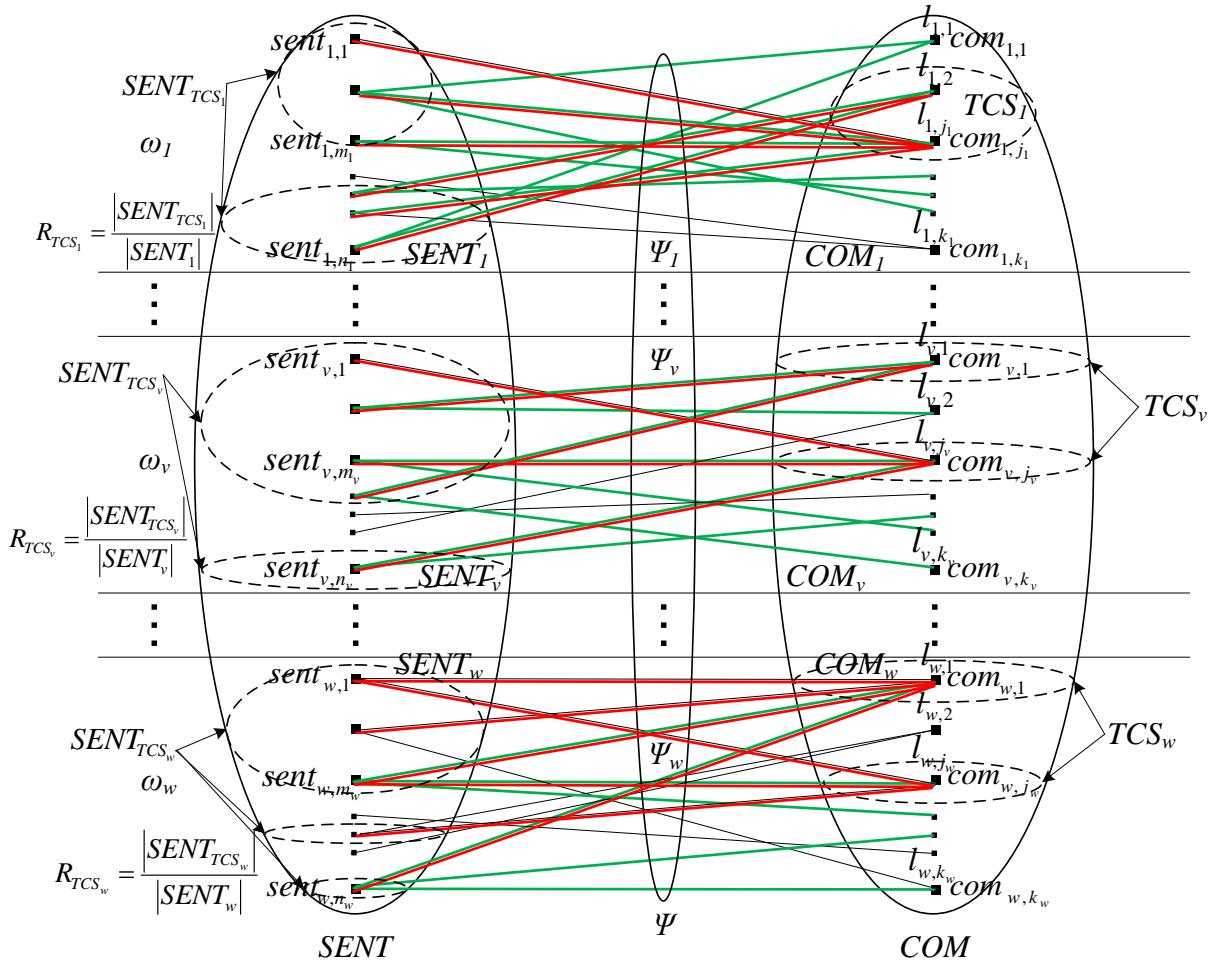


Рисунок 1.30 – Граф задачи нахождения выборки типичных комментариев в ходе анализа результатов обсуждения законопроекта

3. Пусть $p_{v,j_v} = d(com_{v,j_v})$ – количество высказываний в комментарии com_{v,j_v} , определяемое степенью соответствующей вершины графа G , $u_{v,j_v} = f(p_{v,j_v}, l_{v,j_v})$ – свертка показателей p_{v,j_v} и l_{v,j_v} , характеризующая полезность комментария com_{v,j_v} .

Необходимо определить вид свертки u_{v,j_v} и найти такое множество комментариев $UCS = \bigcup_{v=1}^w UCS_v$, $UCS_v = \{com_{v,j_v}\}$, $UCS \subseteq COM$, представляющее собой выборку полезных комментариев, для которого выполняется система требований:

$$\begin{cases} |UCS| \rightarrow \min; \\ \sum_{v=1}^w \sum_{j_v=1}^{h_v} u_{v,j_v} \rightarrow \max; \\ Td = \text{const}, \end{cases} \quad (1.16)$$

где $h_v = |UCS_v|$ – мощность множества UCS_v (рисунок 1.31).

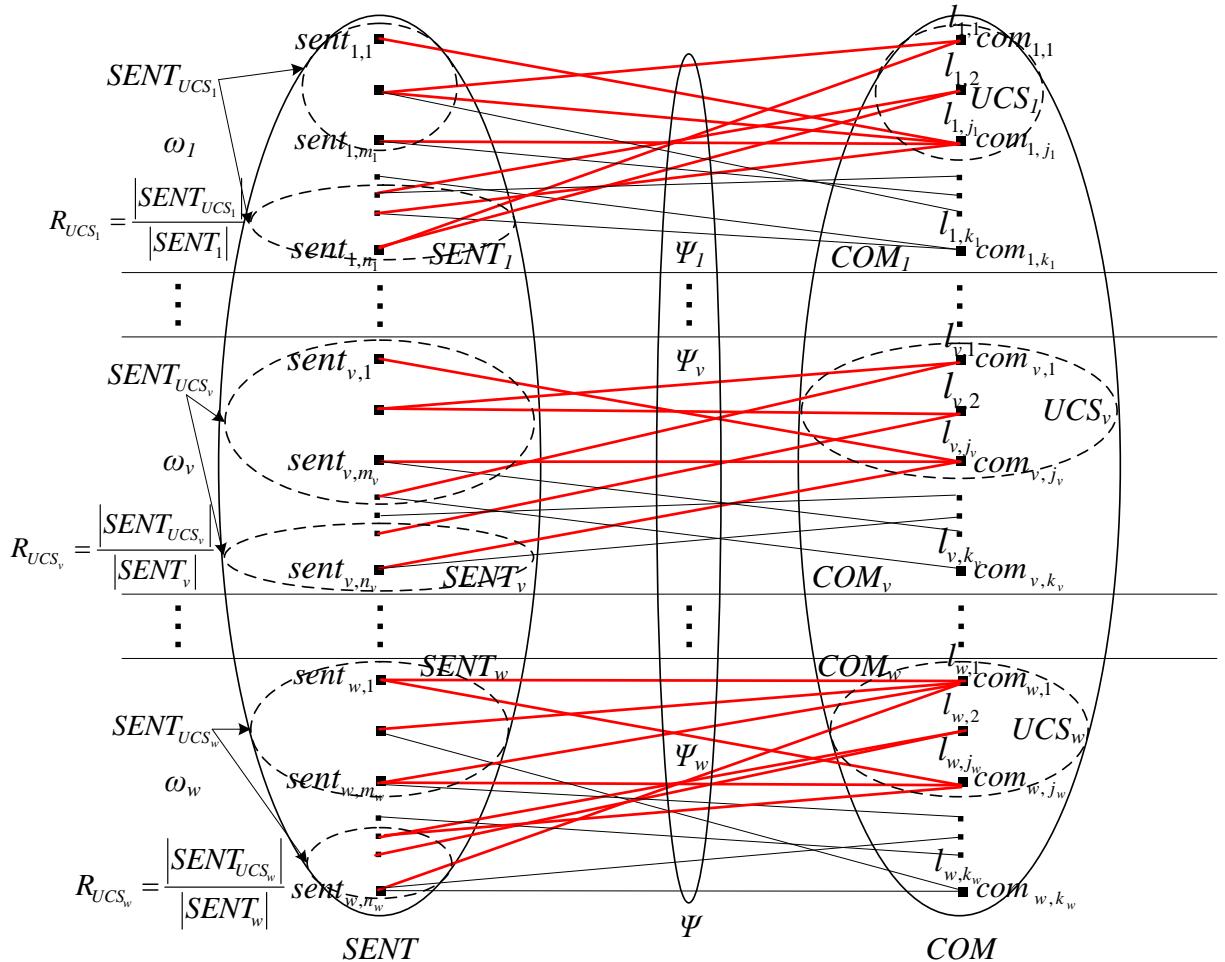


Рисунок 1.31 – Граф задачи нахождения выборки полезных комментариев в ходе анализа результатов обсуждения законопроекта

Выводы по 1 разделу

1. Действующее законодательство устанавливает обязательные требования по проведению и обработке результатов общественного обсуждения в заданные сроки, однако не определяет порядок обработки поступивших предложений.

2. Процесс обработки комментариев экспертами направлен на ознакомление с предложениями по корректировке разрабатываемого законопроекта путем деления исходного массива комментариев на классы, выявление в полученных классах типичных мнений и подкрепление их полезными комментариями.

3. Допустимо теоретико-множественное представление понятия "мнение" в виде кортежа, что позволяет осуществлять выбор метода представления и обработки комментариев.

4. В доступной литературе не выявлено готовых моделей и методов обработки текстовых комментариев к законопроектам, между тем такие модели и методы должны учитывать общие закономерности подходов к обработке мнений: выделение тематических, тональных и стилевых классов, применение обобщения мнений в классах с использованием реферирования, разработка перспективных методов обобщения с использованием выборок.

5. Существующее в исследуемой области теоретическое противоречие заключается в несоответствии между потребностью в моделях аналитической обработки комментариев и отсутствием качественной модели, пригодной для такой обработки. Практическое противоречие состоит в несоответствии между потребностью экспертов в аналитической обработке большого объема комментариев и ограниченными временными и человеческими ресурсами для их обработки.

6. По совокупности характеристик, отражающих существенные свойства текстовых комментариев к законопроектам и процесса их обработки,

предпочтительными являются модели оперативной аналитической обработки текстов.

7. Существующие модели оперативной аналитической обработки текстов недостаточно разработаны применительно к мнениям, а в особенности – к текстовым комментариям, полученным в ходе общественного обсуждения законопроектов.

8. Разрабатываемая модель обработки комментариев должна учитывать следующие основные особенности текстовых комментариев к законопроектам, процедуры их обработки, а также тенденции развития методов обработки мнений:

- возможность оценки интенсивности мнений в классах, сформированных значимыми атрибутами;
- возможность обобщения мнений на основе рефератов и выборок;
- возможность ознакомления с типичными и полезными комментариями в соответствии с выделенным на обработку бюджетом времени.

9. Проведенный анализ позволил выработать рекомендации по построению системы обработки комментариев, которая позволит эксперту понять сущность предложений интернет-пользователей по корректировке обсуждаемого законопроекта и сформировать предложения по использованию комментариев.

10. Произведена формализация задачи исследования, направленная на организацию эффективной обработки комментариев с использованием оперативной аналитической обработки текстов.

11. Для повышения темпа изучения совокупности высказываний в ходе обработки результатов общественного обсуждения законопроектов требуется учитывать многомерность представления мнений и закономерности распределения высказываний в комментариях интернет-пользователей.

2 РАЗРАБОТКА ЭФФЕКТИВНЫХ ПОДХОДОВ К ОБРАБОТКЕ МАССИВА КОММЕНТАРИЕВ НА ОСНОВЕ ОПЕРАТИВНОЙ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ ТЕКСТОВ

2.1 Модель оперативной аналитической обработки комментариев с операциями детализации на основе выявления типичных и полезных текстов

В основе оперативной аналитической обработки [80, 81] лежит формальное представление массива комментариев с использованием гиперкубовой модели $hc : a_1 \times \dots \times a_s \rightarrow a_r$, $s, r \in [1, z]$, $s \neq r$ [82, 83]. Построение гиперкубовой модели осуществляется на основе упорядоченного множества атрибутов A :

$$A = \{a_1, a_2, a_3, a_4, a_5, a_6, a_7\}. \quad (2.1)$$

Общая характеристика атрибутов комментариев приведена в таблице 2.1:

Таблица 2.1 – Общая характеристика атрибутов комментариев

Название атрибута комментария	Условное обозначение	Описание	Множество значений
Глава комментария	a_1	Номер главы законо-проекта, к которой относится комментарий, характеризует тематическую рубрику комментария	$a_1 \in \overline{1, a_{1_{\max}}}$
Статья комментария	a_2	Номер статьи законо-проекта, к которой относится комментарий, характеризует тему комментария	$a_2 \in \overline{1, a_{2_{\max}}}$

Продолжение таблицы 2.1

Название атрибута комментария	Условное обозначение	Описание	Множество значений
Идентификатор комментария	a_3	Уникальный номер, присваиваемый комментарию для его однозначной идентификации в массиве комментариев	$a_3 \in \overline{1, a_{3_{\max}}}$
Функциональный стиль комментария	a_4	Функциональный стиль комментария в соответствии с принятой в языкоznании классификацией	$a_4 \in \{\text{научный, официальный-деловой, газетно-публицистический, художественно-беллетристический, разговорный}\}$
Оценочная тональность комментария	a_5	Характеристика принадлежности комментария к группе поддержки/критики в отношении статей законопроекта	$a_5 \in \{\text{мажорная, минорная}\}$
Эмоциональная тональность комментария	a_6	Характеристика уровня эмоциональности высказываний в отношении статей законопроекта	$a_6 \in \{\text{высокая, средняя, низкая}\}$
Количество комментариев	a_7	Служебный атрибут, необходимый для определения интенсивности мнений в классах при использовании операций количественной агрегации/детализации в гиперкубе	$a_7 \in \overline{0, a_{7_{\max}}}$

На данном множестве экспертным способом устанавливаются функциональные зависимости (таблица 2.2).

Таблица 2.2 – Выявление зависимостей на множестве атрибутов комментариев

Атрибут	Зависимый атрибут	Зависит от атрибута
Глава законопроекта (a_1)	Статья законопроекта (a_2)	–
Статья законопроекта (a_2)	Идентификатор комментария (a_3), количество комментариев (a_7)	Глава законопроекта (a_1)
Идентификатор комментария (a_3)	Функциональный стиль комментария (a_4), оценочная тональность комментария (a_5), эмоциональная тональность комментария (a_6)	Статья законопроекта (a_2)
Функциональный стиль комментария (a_4)	Количество комментариев (a_7)	Идентификатор комментария (a_3)
Оценочная тональность комментария (a_5)	Количество комментариев (a_7)	Идентификатор комментария (a_3)
Эмоциональная тональность комментария (a_6)	Количество комментариев (a_7)	Идентификатор комментария (a_3)
Количество комментариев (a_7)	–	Статья законопроекта (a_2), функциональный стиль комментария (a_4), оценочная тональность комментария (a_5), эмоциональная тональность комментария (a_6)

Представим формально данные функциональные зависимости в виде множества зависимостей:

$$DEP = \{a_1 \rightarrow a_2, a_2 \rightarrow a_3, a_3 \rightarrow a_4, a_3 \rightarrow a_5, a_3 \rightarrow a_6, a_2 a_4 a_5 a_6 \rightarrow a_7\}. \quad (2.2)$$

Полученные зависимости позволяют сформировать измерения $\{D_i\}$ гиперкубовой модели:

$$D_1 = \{a_1 \{a_2 \{a_3\}\}\}; \quad (2.3)$$

$$D_2 = \{a_4\}; \quad (2.4)$$

$$D_3 = \{a_5\}; \quad (2.5)$$

$$D_4 = \{a_6\}, \quad (2.6)$$

и мереу:

$$M = \{a_7\}. \quad (2.7)$$

В итоге гиперкубовая модель представляется выражением:

$$hc : \{a_1 \{a_2 \{a_3\}\}\} \times \{a_4\} \times \{a_5\} \times \{a_6\} \rightarrow a_7. \quad (2.8)$$

Решение задачи (1.15) предлагается осуществлять методом кластеризации [84] комментариев в классах на основе матриц различия, сформированных с использованием матриц подобия [85, 86] с выбором в полученных кластерах типичного комментария.

Выбор матриц подобия в качестве основы для решения задачи обусловлен возможностями мер подобия выявлять содержательное подобие текстов и выступать в качестве прототипа семантической метрики.

Выбор сводных рефератов в качестве типичных представителей кластеров обусловлен свойством методов статистического автоматического реферирования выявлять и включать в реферат высказывания на основе частоты их встречаемости.

Для определения квот в классах используется отношение выделенного директивного времени к максимальному:

$$Q_v = Q_{v_{\max}} \cdot \frac{Td}{Td_{\max}}. \quad (2.9)$$

Полученные квоты определяют число кластеров $i = \overline{1, Q_v}$, на которое производится разбиение множества комментариев класса COM_v :

$$cluster : COM_v \rightarrow \{COM_{v_i}\}, \quad (2.10)$$

где $cluster$ – операция кластеризации, осуществляемая на основе матрицы различия, полученной из матрицы подобия комментариев класса.

В каждом кластере COM_{v_i} производится операция сводного реферирования $svrefer$:

$$\forall COM_{v_i} svrefer : COM_{v_i} \rightarrow ref_{v_i}, \quad (2.11)$$

где ref_{v_i} – сводный реферат [87, 88] комментариев кластера COM_{v_i} , причем режим реферирования должен обеспечивать наилучшее среднее подобие формируемого сводного реферата в кластере комментариев.

Определение типичного комментария в каждом кластере COM_{v_i} производится по принципу медианы посредством операции $typcom$ нахождения комментария $com_{v_{ik}}$, обладающего наилучшим подобием $dist$ с полученным сводным рефератом кластера ref_{v_i} :

$$\forall COM_{v_i} typcom : COM_{v_i} \rightarrow com_{v_{ik}} \mid k = \arg \min dist(com_{v_{ik}}, ref_{v_i}). \quad (2.12)$$

Ознакомление эксперта с выборкой типичных комментариев в классе v осуществляется в пределах квоты Q_v в порядке включения комментариев в выборку с увеличением бюджета времени и ростом числа кластеров COM_{v_i} .

Решение задачи (1.16) предлагается осуществлять путем нахождения максимума функционала обобщенной полезности $f(Q)$ с использованием метода множителей Лагранжа [89, 90] при условии $Td = const$:

$$Q = \arg \max_{Td=const} f(Q), \quad (2.13)$$

где Q – вектор квот для классов комментариев COM_v , оказывающих значимое влияние на обобщенный показатель полезности u_{v, j_v} :

$$u_{v, j_v} = \alpha_p \cdot \frac{p_{v, j_v}}{p_{\max}} + \alpha_l \cdot \frac{l_{v, j_v}}{l_{\max}}, \quad (2.14)$$

где α_p , α_l – коэффициенты значимости, подбираемые с учетом распределений показателей свертки [91].

Применение метода множителей Лагранжа обусловлено возможностью построения аналитических выражений для убывающих функций предельной полезности комментариев в каждом классе, наличием временной стоимости обработки комментариев, связанной с темпом чтения и длиной комментария, и возможностью сформировать уравнение бюджетного ограничения (в данной задаче стоимость измеряется в единицах времени).

Поскольку кривая, характеризующая убывание обобщенного показателя в классах COM_v (кривая предельной полезности), со средней достоверностью 0,9 аппроксимируется логарифмической функцией вида $g_v(Q_v) = A_v - B_v \cdot \ln(Q_v)$, функционал $f(Q)$ принимает вид:

$$f(Q) = \sum_{v=1}^w f(Q_v); \quad (2.15)$$

$$f(Q_v) = \int_{Q_v} A_v - B_v \cdot \ln(Q_v) dQ_v; \quad (2.16)$$

$$f(Q) = \sum_{v=1}^w [Q_v \cdot (A_v + B_v) - B_v \cdot Q_v \cdot \ln(Q_v)]. \quad (2.17)$$

Задание директивного времени Td осуществляется при условии:

$$0 < Td \leq Td_{\max}, \quad (2.18)$$

причем общее уравнение бюджетного ограничения (по времени) примет вид:

$$Td = \sum_{v=1}^w L_v \cdot T1 \cdot Q_v; \quad (2.19)$$

$$L_v = \frac{\sum_{j_v=1}^{k_v} W_{v,j_v}}{|COM_v|}, \quad (2.20)$$

где L_v – среднее количество слов в комментарии класса v , $T1$ – среднее время чтения одного слова, Q_v – число комментариев выборки класса v , W_{v,j_v} – количество слов в комментарии j_v класса v , соответственно уравнение для максимального директивного времени, необходимого для изучения всех комментариев в классах, имеет вид:

$$Td_{\max} = \sum_{v=1}^w L_v \cdot T1 \cdot Q_{v_{\max}}. \quad (2.21)$$

В реальных условиях естественное ограничение накладывается и на количество комментариев в выборках классов Q_v :

$$1 \leq Q_v \leq Q_{v_{\max}}. \quad (2.22)$$

Из теории полезности [92, 93] известно, что функционал в уравнении (2.17) принимает максимальное значение, если каждая из его частных производных и уравнение бюджетного ограничения равны нулю. При этом для решения системы уравнений вводится вспомогательная переменная, известная как множитель Лагранжа λ , и производятся следующие преобразования: уравнение бюджетного ограничения приравнивается к нулю, умножается на λ , затем полученный результат складывается с уравнением (2.17). В результате выполнения данных действий получается выражение:

$$f(Q) = \sum_{v=1}^w [Q_v \cdot (A_v + B_v) - B_v \cdot Q_v \cdot \ln(Q_v)] + \lambda \cdot (Td - \sum_{v=1}^w L_v \cdot T1 \cdot Q_v). \quad (2.23)$$

Взяв частные производные от уравнения (2.23), получаем:

$$\frac{\partial f(Q)}{\partial Q_v} = A_v - B_v \cdot \ln(Q_v) - \lambda \cdot L_v \cdot T1 = 0; \quad (2.24)$$

$$\frac{\partial f(Q)}{\partial \lambda} = Td - \sum_{v=1}^w L_v \cdot T1 \cdot Q_v = 0. \quad (2.25)$$

Уравнение (2.25) говорит о максимизации полезности при израсходовании всего имеющегося бюджета времени. Разделив уравнение (2.24) на $L_v \cdot T1$, получаем:

$$\frac{A_v - B_v \cdot \ln(Q_v)}{L_v \cdot T1} = \lambda. \quad (2.26)$$

В итоге задача (2.13) с учетом выражений (2.22), (2.25), (2.26) сводится к системе:

$$\left\{ \begin{array}{l} Td = \sum_{v=1}^w L_v \cdot T1 \cdot Q_v; \\ \frac{A_1 - B_1 \cdot \ln(Q_1)}{L_1 \cdot T1} = \frac{A_2 - B_2 \cdot \ln(Q_2)}{L_2 \cdot T1}; \\ \frac{A_2 - B_2 \cdot \ln(Q_2)}{L_2 \cdot T1} = \frac{A_3 - B_3 \cdot \ln(Q_3)}{L_3 \cdot T1}; \\ \dots \\ \frac{A_{v-1} - B_{v-1} \cdot \ln(Q_{v-1})}{L_{v-1} \cdot T1} = \frac{A_v - B_v \cdot \ln(Q_v)}{L_v \cdot T1}; \\ \dots \\ \frac{A_{w-1} - B_{w-1} \cdot \ln(Q_{w-1})}{L_{w-1} \cdot T1} = \frac{A_w - B_w \cdot \ln(Q_w)}{L_w \cdot T1}; \\ 1 \leq Q_1 \leq Q_{1_{\max}}; \\ 1 \leq Q_2 \leq Q_{2_{\max}}; \\ \dots \\ 1 \leq Q_v \leq Q_{v_{\max}}; \\ \dots \\ 1 \leq Q_{w-1} \leq Q_{w-1_{\max}}; \\ 1 \leq Q_w \leq Q_{w_{\max}}, \end{array} \right. \quad (2.27)$$

решение которой осуществляется численными методами с использованием математических пакетов прикладных программ и определяет вектор квот:

$$Q = (Q_1, Q_2, \dots, Q_v, \dots, Q_w). \quad (2.28)$$

Ознакомление эксперта с выборкой полезных комментариев в классе v осуществляется в пределах квоты Q_v в порядке убывания обобщенного показателя полезности комментария u_{v, j_v} .

Предложенная модель оперативной аналитической обработки комментариев с операциями детализации на основе выявления типичных и полезных текстов позволяет:

- осуществлять анализ массива комментариев как массива многомерных данных с использованием стандартных OLAP-операций и устанавливать порядок обработки классов комментариев (обращения к ячейкам гиперкуба) с учетом интенсивности мнений;
- осуществлять операцию погружения в ходе анализа массива комментариев с использованием качественной детализации на основе выборки типичных комментариев, учитывающей содержательное подобие текстов комментариев, сходство со сводным рефератом кластера и выделенное директивное время на обработку комментариев для их приоритетного отбора;
- осуществлять операцию погружения в ходе анализа массива комментариев с использованием качественной детализации на основе выборки полезных комментариев, учитывающей неравномерное распределение конструктивных высказываний в комментариях, уровень поддержки интернет-пользователей и выделенное директивное время на обработку комментариев для их приоритетного отбора.

2.2 Алгоритм формирования выборки типичных комментариев

Алгоритм решения задачи (1.15) представлен на рисунке 2.1.



Рисунок 2.1 – Алгоритм формирования выборки типичных комментариев

Сущность данного алгоритма заключается в разбиении, в соответствии с заданным бюджетом времени, исходного множества комментариев в классах (ячейках гиперкуба) на кластеры на основе матриц подобия и определении множества комментариев, по одному в каждом кластере, обладающих наилучшим сходством с типичными представителями кластеров комментариев, в качестве которых выступают сводные рефераты кластеров [94, 95].

В ходе экспериментов установлено, что сводные рефераты обладают наилучшим средним подобием в классе комментариев (таблицы 2.3 и 2.4). Указанное свойство было положено в основу процедуры определения типичных комментариев.

Таблица 2.3 – Свойство наилучшего среднего подобия сводных рефератов комментариев различных классов на примере законопроекта "Об образовании" (модель подобия – марковская [96], рефераты получены методом сканирующих статистик [97])

Текст (класс "34-ГП-Н-Ми")	Ср. коэф. подобия	Текст (класс "11-ОД-Н-Ми")	Ср. коэф. подобия	Текст (класс "1-ОД-Н-Ми")	Ср. коэф. подобия
edu 34-2.txt	0,62089	edu 11-3.txt	0,163084	edu 1-4.txt	0,323633
edu 34-3.txt	0,626469	edu 11-4.txt	0,163084	edu 1-3.txt	0,324475
edu 34-4.txt	0,62749	edu 11-1.txt	0,163153	edu 1-1.txt	0,326142
edu 34-1.txt	0,629656	edu 11-2.txt	0,163153	edu 1-2.txt	0,3335
4166	0,986917	2343	0,226228	7732	0,600645
520	1,071068	2823	0,289161	5520	0,658391
458	1,078332	2346	0,306828	6276	0,667973
4572	1,121677	513	0,502133	918	0,674645
4106	1,133296	7754	0,503078	5340	0,726827
3383	1,153315	3616	0,508561	5488	1,032082
..

Таблица 2.4 – Свойство наилучшего среднего подобия сводных рефератов комментариев различных классов на примере законопроекта "О полиции" (модель подобия – марковская, рефераты получены методом сканирующих статистик)

Текст (класс "3-12-Од-В-Ми")	Ср. коэф. подобия	Текст (класс "7-42-Н-В-Ми")	Ср. коэф. подобия	Текст (класс "6-34-Н-Н-Ми")	Ср. коэф. подобия
pol 12-4.txt	0,401729	pol 42-3.txt	0,323454	pol 34-1.txt	0,413644
pol 12-1.txt	0,408188	pol 42-4.txt	0,323804	pol 34-3.txt	0,423178
pol 12-3.txt	0,416202	pol 42-2.txt	0,327507	pol 34-2.txt	0,4248
pol 12-2.txt	0,422409	pol 42-1.txt	0,567832	pol 34-4.txt	0,441756
11011	0,783439	2839	0,979496	2505	0,555188
10503	0,831565	16281	1,090419	15793	0,555813
10037	0,837542	15430	1,241689	6239	0,582438
266	0,960356	1779	1,32573	6205	0,613938
151	1,000195	18525	1,330104	14322	0,633663
18349	1,134112	5639	1,334311	5601	1,042525
..

Тексты "edu XX-Y" и "pol XX-Y" в таблицах 2.3 и 2.4 представляют собой сводные рефераты соответствующих классов, при этом "XX" кодирует номер статьи, а "Y" – режим реферирования. Для получения сводных рефератов применялись режимы автоматического реферирования с использованием метода сканирующих статистик, представленные в таблице 2.5.

Таблица 2.5 – Режимы функционирования алгоритма автоматического реферирования методом сканирующих статистик

№ режима	Окно поиска	Отношение ключевые/все слова	Количество предложений
1	6	0,0009	8
2	7	0,0006	11
3	8	0,0003	13
4	9	0,0001	17

В ходе исследования было установлено, что наилучшее среднее подобие по каждому из законопроектов показали рефераты, полученные в режиме автоматического реферирования № 3.

Рассмотрим пошагово выполнение алгоритма формирования выборки типичных комментариев.

На шаге 1 алгоритма задается бюджет времени Td с учетом ограничения $0 \leq Td \leq Td_{\max}$, производится выбор модели подобия и метода реферирования.

На шаге 2, исходя из заданного бюджета времени и мощности класса $Q_{v_{\max}} = |COM_v|$, определяется мощность выборки типичных комментариев. Так, для класса $v="7-42-Н-В-Ми"$ при $Td = 0,5 \cdot Td_{\max}$ мощность выборки будет составлять:

$$Q_v = 28 \cdot 0,5 = 14 \text{ (комментариев).} \quad (2.29)$$

На шаге 3 для выбранного класса комментариев строится матрица подобия M_v , на основе которой, в зависимости от используемой модели подобия, формируется матрица различий (таблица 2.6).

Таблица 2.6 – Фрагмент матрицы различий (модель подобия – марковская, класс "7-42-Н-В-Ми" законопроекта "О полиции")

Идент. комм.	19074	359	1658	1779	2715	2839	3875	...
19074	0	2,7175	3,0272	3,0839	3,1179	2,5739	4,2059	...
359	2,6959	0	0,9163	1,0651	0,3536	0,156	0,8233	...
1658	2,9231	0,9522	0	1,1987	1,3682	0,1889	2,7109	...
1779	2,8662	1,0529	1,206	0	0,2912	0,926	2,8638	...
2715	3,0401	0,3536	1,384	0,2912	0	1,1362	3,1515	...
2839	2,6374	0,156	0,1889	0,901	1,1147	0	0,4452	...
3875	4,1056	0,8233	2,5657	2,6757	2,8837	0,4452	0	...
...

Особенностью матрицы различий является соответствие меньших значений коэффициентов более близким по содержанию комментариям. Если коэффициенты в матрице подобия не отвечают данному требованию и интерпретируются с противоположным смыслом, необходимо осуществить переход к матрице различий. Перевод может быть осуществлен по формуле:

$$D_r = D_{p_{\max}} - D_p, \quad (2.30)$$

где D_r – коэффициент различия, D_p – коэффициент подобия, $D_{p_{\max}}$ – максимальное значение коэффициента в матрице подобия.

На шаге 4 к полученной матрице различий применяется метод иерархического объединения для кластеризации комментариев на требуемое количество кластеров Q_v в соответствии с заданным бюджетом времени Td (рисунок 2.2).

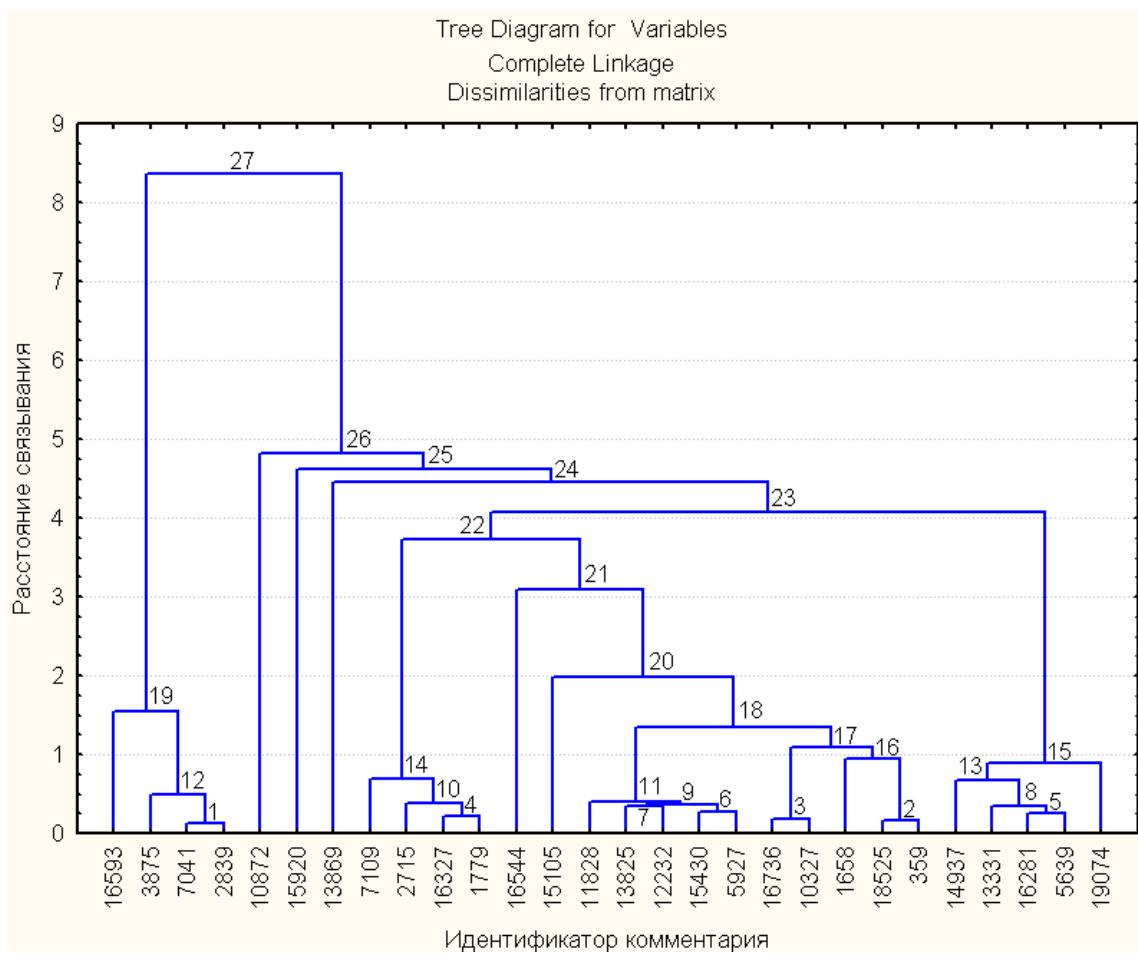


Рисунок 2.2 – Результат кластеризации комментариев (метод кластеризации – иерархическое объединение, класс "7-42-Н-В-Ми" законопроекта "О полиции")

Шаги 5, 10 и проверка условия 11 обеспечивают итерацию шагов 6–9 по всем кластерам, сформированным на шаге 4.

На шаге 6 в полученных кластерах строятся сводные рефераты комментариев и включаются в массив соответствующего кластера.

На шаге 7 формируются матрицы подобия в расширенных полученными сводными рефератами кластерах (таблица 2.7).

Таблица 2.7 – Матрица подобия 15 кластера (модель подобия – марковская, класс "7-42-Н-В-Ми" законопроекта "О полиции")

Текст	5639.txt	13331.txt	14937.txt	16281.txt	19074.txt	all15ref3.txt
5639.txt	0	0,2624	0,6736	0,2474	0,8242	0,2838
13331.txt	0,2624	0	0,6804	0,3556	0,8036	0,0591
14937.txt	0,6736	0,6804	0	0,674	0,8996	2,5486
16281.txt	0,2474	0,3556	0,674	0	0,8227	0,0834
19074.txt	0,8242	0,8036	0,8996	0,8227	0	2,9657
all15ref3.txt	0,2838	0,0591	2,7037	0,0834	3,1937	0

На шаге 8 в расширенных кластерах производится ранжирование комментариев по подобию сводному реферату кластера.

На шаге 9 в полученных ранжированных списках определяются комментарии, обладающие наилучшим сходством со сводным рефератом своего кластера. Так, для 15 (по порядку создания) кластера класса "7-42-Н-В-Ми" законопроекта "О полиции" данному критерию соответствует комментарий с идентификатором "13331", который получает статус "типичный комментарий" в своем кластере (таблица 2.8).

Таблица 2.8 – Выбор типичного комментария в 15 кластере (модель подобия – марковская, класс "7-42-Н-В-Ми" законопроекта "О полиции")

Текст	5639.txt	13331.txt	14937.txt	16281.txt	19074.txt	all15ref3.txt
all15ref3.txt	0,2838	0,0591	2,7037	0,0834	3,1937	0
13331.txt	0,2624	0	0,6804	0,3556	0,8036	0,0591
16281.txt	0,2474	0,3556	0,674	0	0,8227	0,0834
5639.txt	0	0,2624	0,6736	0,2474	0,8242	0,2838
14937.txt	0,6736	0,6804	0	0,674	0,8996	2,5486
19074.txt	0,8242	0,8036	0,8996	0,8227	0	2,9657

На шаге 12 совокупность отобранных в каждом из кластеров комментариев составляет выборку типичных комментариев указанного экспертом класса, которые подлежат изучению в порядке включения в выборку при увеличении числа кластеров (таблица 2.9).

Таблица 2.9 – Формирование выборки типичных комментариев при увеличении числа кластеров (модель подобия – марковская, класс "7-42-Н-В-Ми" законопроекта "О полиции")

К-во комм.	Номер кластера																											Поряд. вкл/оч.		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27			
1	2839	359	10327	1779	5639	5927	12232	5639	5927	1779	5927	2839	5639	1779	19074	359	359	2839	359	359	359	19074	19074	19074	19074	19074	15			
2	7041	18525	16736	16327	16281	15430	13825	16281	15430	16327	15430	7041	16281	16327	5639	18525	18525	18525	7041	18525	18525	18525	5639	5639	5639	5639	5639	24		
3								13331	12232	2715	12232	3875	13331	2715	16281	1658	1658	1658	3875	1658	1658	1658	16281	16281	16281	16281	16281	23		
4										13825		13825		14937	7109	13331		10327	10327	16593	10327	10327	10327	13331	13331	13331	13331	13331	6	
5													11828				14937		16736	16736	16736	16736	14937	14937	14937	14937	14937	17		
6																	5927		5927	5927	5927	359	359	359	359	359	27			
7																		15430	15430	15430	15430	18525	18525	18525	18525	18525	14			
8																		12232	12232	12232	12232	1658	1658	1658	1658	1658	12			
9																		13825	13825	13825	13825	10327	10327	10327	10327	10327	26			
10																		11828		11828	11828	11828	16736	16736	16736	16736	16736	13		
11																			15105	15105	15105	15105	5927	5927	5927	5927	5927	21		
12																			16544	16544	15430	15430	15430	15430	15430	15430	15430	7		
13																			1779	12232	12232	12232	12232	12232	12232	12232	12232	19		
14																			16327	13825	13825	13825	13825	13825	13825	13825	13825	22		
15																			2715	11828	11828	11828	11828	11828	11828	11828	11828	11		
16																			7109	15105	15105	15105	15105	15105	15105	15105	15105	9		
17																				16544	16544	16544	16544	16544	16544	16544	16544	8		
18																			1779	1779	1779	1779	1779	1779	1779	1779	2			
19																			16327	16327	16327	16327	16327	16327	16327	16327	25			
20																			2715	2715	2715	2715	2715	2715	2715	2715	20			
21																			7109	7109	7109	7109	7109	7109	7109	7109	16			
22																				13869	13869	13869	13869	13869	13869	13869	13869	5		
23																				15920	15920	15920	15920	15920	15920	15920	15920	4		
24																					10872	10872	10872	10872	10872	10872	10872	10872	3	
25																					2839								1	
26																											7041	28		
27																												3875	18	
28																													16593	10

Например, для класса "7-42-Н-В-Ми" законопроекта "О полиции" при квоте $Q_v = 7$ для $Td = 0,25 \cdot Td_{\max}$ ознакомление с комментариями в соответствии с таблицей 2.18 следует осуществлять следующем порядке:

$$2839 \rightarrow 1779 \rightarrow 10872 \rightarrow 15920 \rightarrow 13869 \rightarrow 13331 \rightarrow 15430. \quad (2.31)$$

Предложенный алгоритм формирования выборки типичных комментариев позволяет:

– находить такое множество комментариев $TCS = \{com_{v, j_v}\}$,

$TCS \subseteq COM$, представляющее собой выборку типичных комментариев, для которого выполняется система требований:

$$\begin{cases} |TCS| \rightarrow \min; \\ \sum_{v=1}^w \sum_{j_v=1}^{g_v} y_{v, j_v} \rightarrow \max; \\ Td = const; \end{cases} \quad (2.32)$$

– учесть содержательное подобие комментариев и сходство с типичным представителем мнения при формировании выборки типичных комментариев на основе кластеризации с использованием матриц подобия с последующим определением в кластерах комментариев, обладающих наилучшим

сходством со сводным рефератом кластера;

– аккумулировать в комментариях выборки типичные высказывания в наиболее удаленных по содержанию кластерах подобных комментариев;

– обеспечить получение данных для проведения качественной детализации на основе формирования выборки типичных комментариев в ходе оперативной аналитической обработки комментариев.

2.3 Алгоритм формирования выборки полезных комментариев

Алгоритм решения задачи (1.16) представлен на рисунке 2.3.

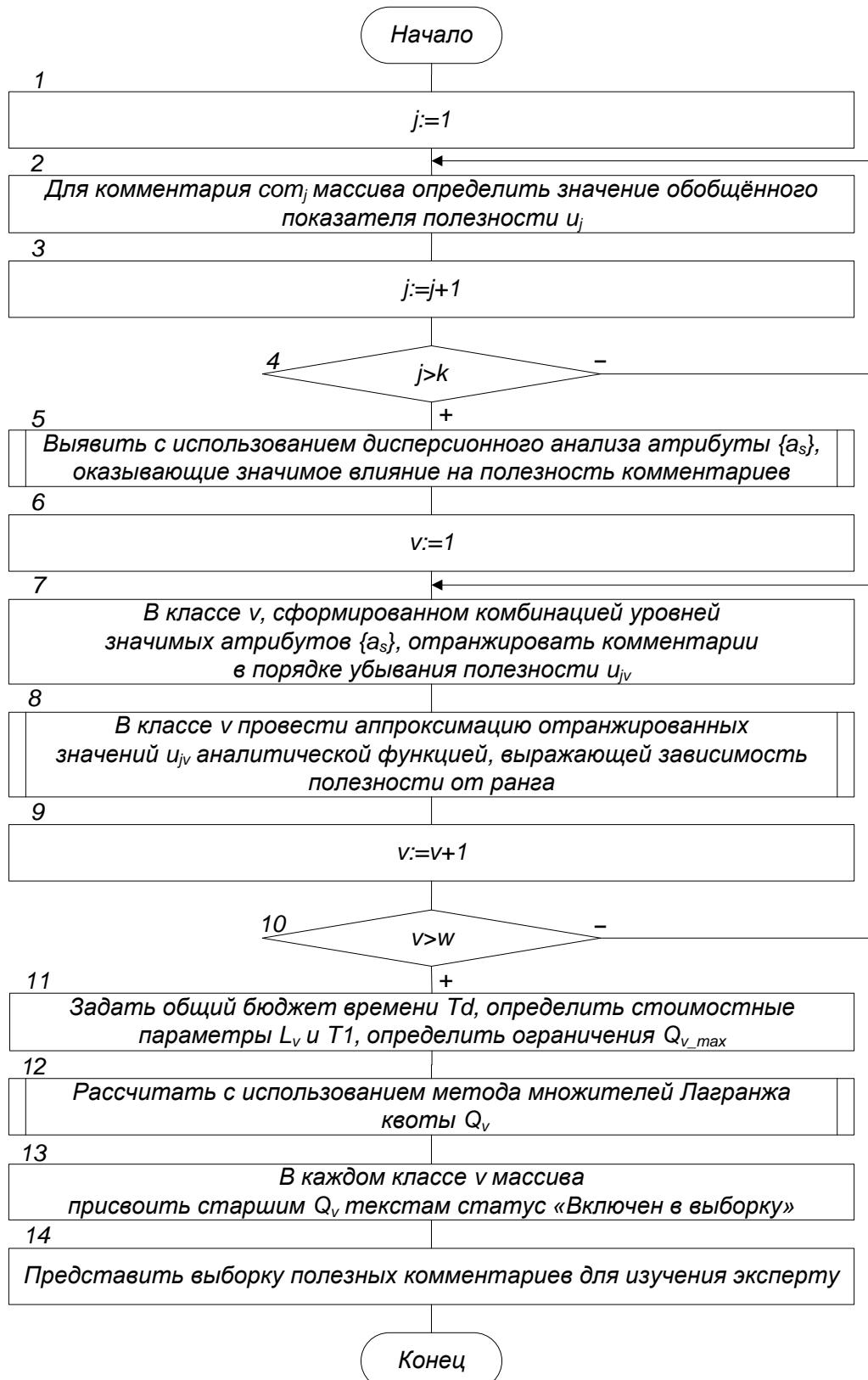


Рисунок 2.3 – Алгоритм формирования выборки полезных комментариев

Сущность данного алгоритма заключается в определении обобщенного показателя полезности комментариев, выявлении значимых атрибутов комментариев, установлении параметров аналитических функций предельной полезности комментариев в классах, образованных значимыми атрибутами, определении стоимостных параметров, задании бюджета времени и решении оптимизационной задачи с использованием метода множителей Лагранжа [93].

Рассмотрим пошагово выполнение данного алгоритма.

Шаги 1, 3 и проверка условия 4 обеспечивают итерацию шага 2 по всем комментариям.

На шаге 2 производится расчет обобщенного показателя полезности.

Исследование материалов обзора к законопроекту "О полиции" позволило выявить две основные составляющие полезности,ываемые экспертами при отборе мнений (рисунок 2.4). Оценка данных составляющих может быть проведена на основе анализа числа "лайков" и атрибутивного анализа.

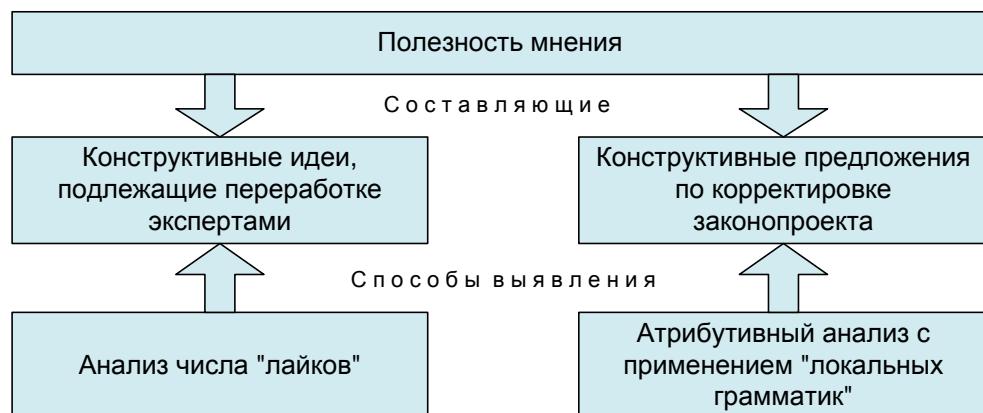


Рисунок 2.4 – Составляющие полезности мнения и способы их выявления

Обобщенный показатель полезности каждого комментария целесообразно рассчитывать с применением аддитивной свертки числа голосов, отданных за комментарий, ("лайков") и числа конструктивных предложений, полученных с использованием атрибутивного анализа, с учетом распределений данных параметров:

$$u_{v, j_v} = \alpha_p \cdot \frac{p_{v, j_v}}{p_{\max}} + \alpha_l \cdot \frac{l_{v, j_v}}{l_{\max}}. \quad (2.33)$$

Определение коэффициентов значимости α_p и α_l предлагается выполнять экспериментально с использованием распределений комментариев по соответствующим параметрам. Установлено, что они представлены экспоненциальным распределением:

$$h(x, \lambda) = \lambda \cdot e^{-\lambda \cdot x} \quad (2.34)$$

с эмпирически подобранными значениями его параметров (рисунок 2.5):

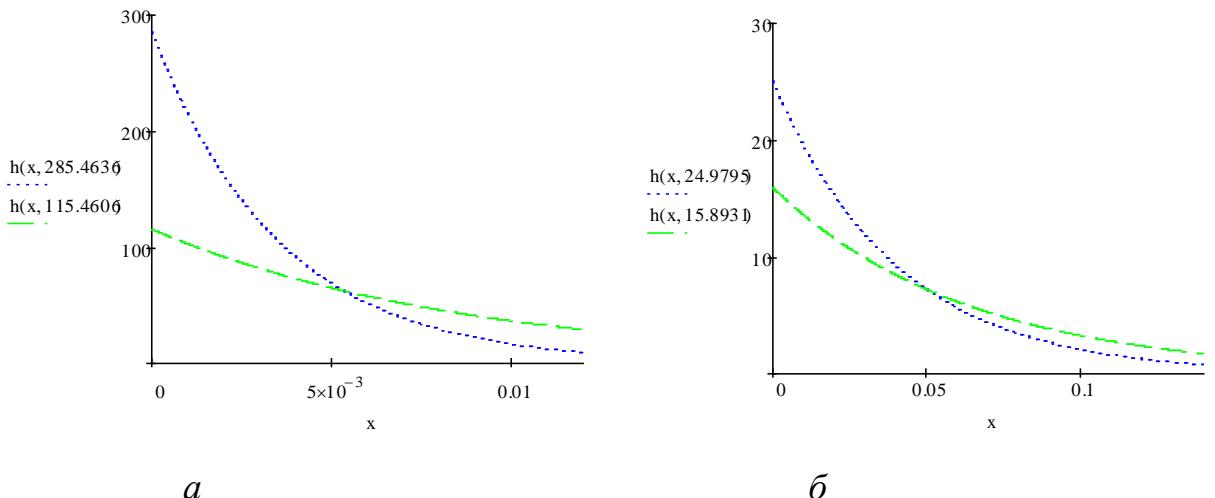


Рисунок 2.5 – Распределение комментариев: *а* – по числу лайков и *б* – конструктивных предложений (после нормирования, без масштаба) в исходном массиве и обзоре к законопроекту "О полиции"

Выбор аддитивной свертки обусловлен возможностью сравнивать разнородные свойства комментариев по абсолютным значениям, получать отличное от нуля значение обобщенного показателя полезности при нулевых значениях одного из частных показателей, а также свойством данного вида

свертки давать большее значение обобщенного показателя для объектов с большими значениями частных параметров.

В целях обеспечения одинаковой чувствительности обобщенного показателя к изменениям составляющих, при расчете α_p и α_l потребовалось сопоставление сходных характеристик распределений по нормированным параметрам p и l (таблицы 2.10 и 2.11).

1. Сопоставление для законопроекта "О полиции".

Для нормированного количества голосов, отданных за комментарий, ("лайков") функция плотности распределения имеет вид:

$$h(x, \lambda) = 285,4636 \cdot e^{-285,4636 \cdot x}. \quad (2.35)$$

Для нормированного количества конструктивных предложений функция плотности распределения имеет вид:

$$h(x, \lambda) = 24,9795 \cdot e^{-24,9795 \cdot x}. \quad (2.36)$$

Составим таблицу накопления вероятностей (таблица 2.10).

Таблица 2.10 – Накопление вероятности при линейном изменении нормированного числа "лайков" и предложений (законопроект "О полиции")

Наблюдение	1	2	3	4	5	6	7
"Лайков" (нормир.)	0,003	0,006	0,009	0,012	0,015	0,018	0,021
Предложений (нормир.)	0,035	0,07	0,105	0,14	0,175	0,21	0,245
$F_{\text{лайков}}$	0,575	0,82	0,923	0,967	0,986	0,994	0,998
$F_{\text{предложений}}$	0,583	0,826	0,927	0,97	0,987	0,995	0,998

Из таблицы видно, что плотности распределения параметров аналогичны по форме, однако накопление вероятностей происходит в разных областях. Проводя расчеты на основе выражения (2.33), данных из таблицы 2.10 и при выполнении условий аддитивной свертки:

$$\begin{cases} 0,018 \cdot \alpha_l - 0,21 \cdot \alpha_p = 0; \\ \alpha_l + \alpha_p = 1, \end{cases} \quad (2.37)$$

получаем $\alpha_l \approx 0,921$, $\alpha_p \approx 0,079$. Распределение комментариев по обобщенному показателю полезности и для законопроекта "О полиции" представлено на рисунке 2.6.

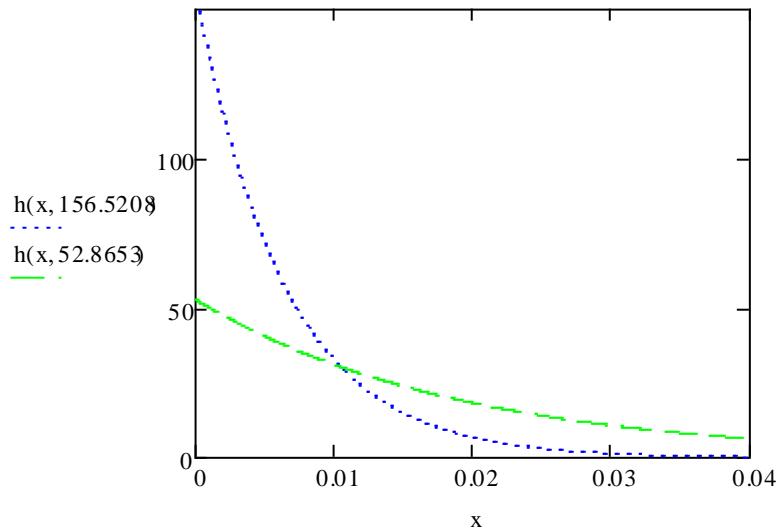


Рисунок 2.6 – Распределение комментариев по обобщенному показателю полезности в исходном массиве и в аналитическом обзоре

Из графика видно, что при составлении аналитического обзора эксперты отбирают более полезные по предложенному обобщенному показателю комментарии, что указывает на пригодность данного показателя для оценки полезности комментариев.

2. Сопоставление для законопроекта "Об образовании".

Проведем аналогичные расчеты для комментариев к законопроекту "Об образовании". Для нормированного количества голосов, отданных за комментарий, ("лайков") распределение имеет вид:

$$h(x) = 148,3521 \cdot e^{-148,3521 \cdot x}. \quad (2.38)$$

Для нормированного количества конструктивных предложений распределение имеет вид:

$$h(x) = 17,0218 \cdot e^{-17,0218 \cdot x}. \quad (2.39)$$

Составим таблицу накопления вероятностей (таблица 2.11).

Таблица 2.11 – Накопление вероятности при линейном изменении нормированного числа "лайков" и предложений (законопроект "Об образовании")

Наблюдение	1	2	3	4	5	6	7
"Лайков" (нормир.)	0,005(6)	0,011(3)	0,017	0,022(6)	0,028(3)	0,034	0,039(6)
Предложений (нормир.)	0,05	0,1	0,15	0,2	0,25	0,3	0,35
$F_{\text{лайков}}$	0,571	0,813	0,92	0,966	0,985	0,994	0,997
$F_{\text{предложений}}$	0,573	0,818	0,922	0,967	0,986	0,994	0,997

Из таблицы видно, что плотности распределения параметров также аналогичны по форме, при этом накопление вероятностей происходит в разных областях. Проводя расчеты на основе выражения (2.33), данных из таблицы 2.11 и при выполнении условий аддитивной свертки:

$$\begin{cases} 0,034 \cdot \alpha_l - 0,3 \cdot \alpha_p = 0; \\ \alpha_l + \alpha_p = 1, \end{cases} \quad (2.40)$$

получаем $\alpha_l \approx 0,8982$, $\alpha_p \approx 0,1018$.

Полученные при решении систем (2.37) и (2.40) решения позволяют выполнить шаг 2 алгоритма формирования выборки полезных комментариев и произвести расчет обобщенного показателя полезности в соответствии с выражением (2.33). Фрагмент результатов расчета для законопроектов "О полиции" и "Об образовании" представлен в таблицах 2.12 и 2.13.

Таблица 2.12 – Результаты расчета обобщенного показателя полезности комментариев к законопроекту "О полиции" (фрагмент)

Идентификатор комментария	Голосов ("лайков")	Конструктивных предложений	Голосов нормир.	Предложений нормир.	Свертка
...
11099	1	3	0,00108	0,157895	0,013468
11100	1	1	0,00108	0,052632	0,005152
11101	0	4	0	0,210526	0,016632
11102	0	0	0	0	0
11103	0	1	0	0,052632	0,004158
11104	1	1	0,00108	0,052632	0,005152
11105	0	1	0	0,052632	0,004158
11106	0	1	0	0,052632	0,004158
11107	1	0	0,00108	0	0,000995
11108	0	2	0	0,105263	0,008316
...

Таблица 2.13 – Результаты расчета обобщенного показателя полезности комментариев к законопроекту "Об образовании" (фрагмент)

Идентификатор комментария	Голосов ("лайков")	Конструктивных предложений	Голосов нормир.	Предложений нормир.	Свертка
...
7256	5	0	0,002974	0	0,002672
7257	4	2	0,00238	0,166667	0,019104
7258	31	1	0,018441	0,083333	0,025047
7259	1	0	0,000595	0	0,000534
7260	16	2	0,009518	0,166667	0,025516
7261	23	1	0,013682	0,083333	0,020773
7262	2	1	0,00119	0,083333	0,009552
7263	3	1	0,001785	0,083333	0,010086
7264	76	0	0,045211	0	0,040609
7265	1	1	0,000595	0,083333	0,009018
...

На шаге 5 атрибуты, оказывающие значимое влияние на обобщенный показатель полезности комментариев, целесообразно определять с использованием дисперсионного анализа [98, 99].

Результаты дисперсионного анализа, выполненного в пакете Statistica 6, на примере комментариев к законопроекту "О полиции" приведены в таблице 2.14 [100, 101].

Таблица 2.14 – Результаты дисперсионного анализа для комментариев к законопроекту "О полиции"

Эффект	SS	Степ. свободы	MS	F	p
Статья	0,064	56	0,001	2,75	0,000*
Стиль	0,15	4	0,037	89,6	0,000*
Тон(эмоц)	0,001	2	0	0,77	0,462
Тон(оцен)	0,003	1	0,003	6,56	0,010*

Графики, описывающие влияние атрибутов на обобщенный показатель полезности комментариев к законопроекту "О полиции", приведены на рисунках 2.7–2.10.

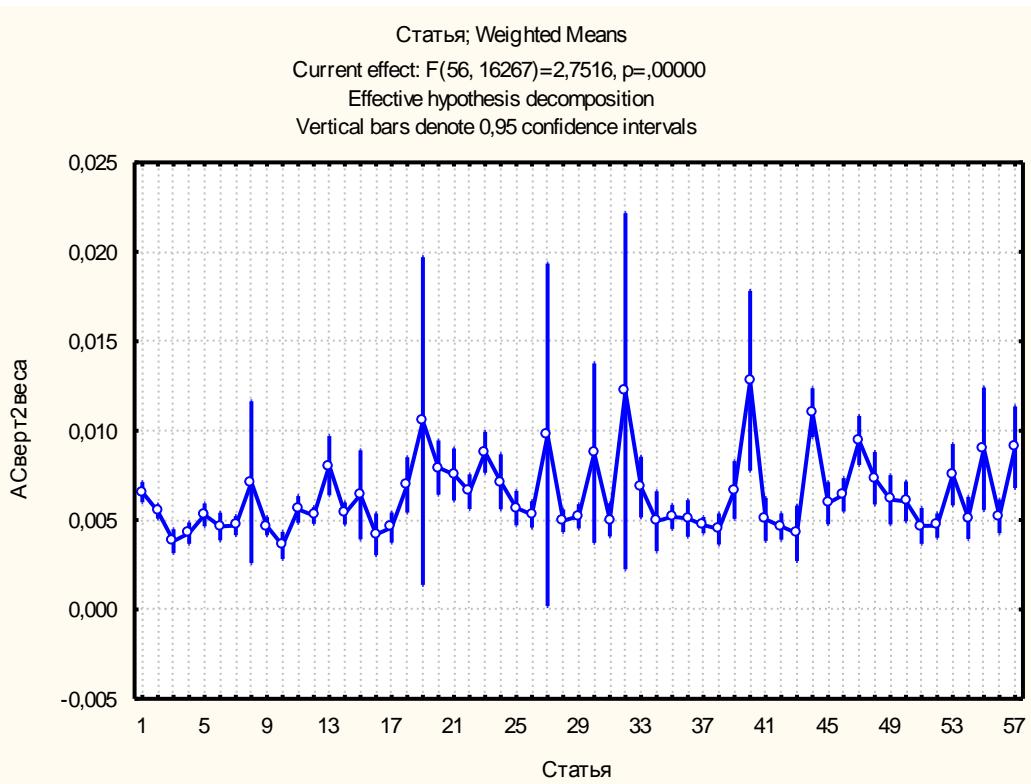


Рисунок 2.7 – Влияние фактора "Статья" на обобщенный показатель полезности комментариев к законопроекту "О полиции"

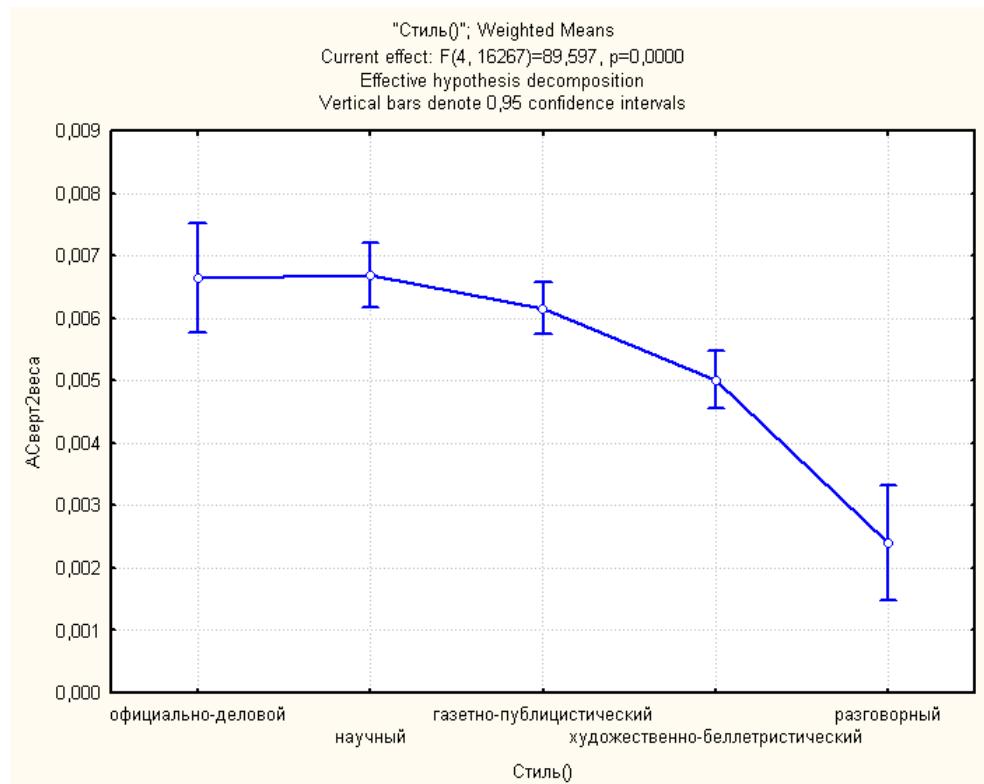


Рисунок 2.8 – Влияние фактора "Стиль" на обобщенный показатель полезности комментариев к законопроекту "О полиции"

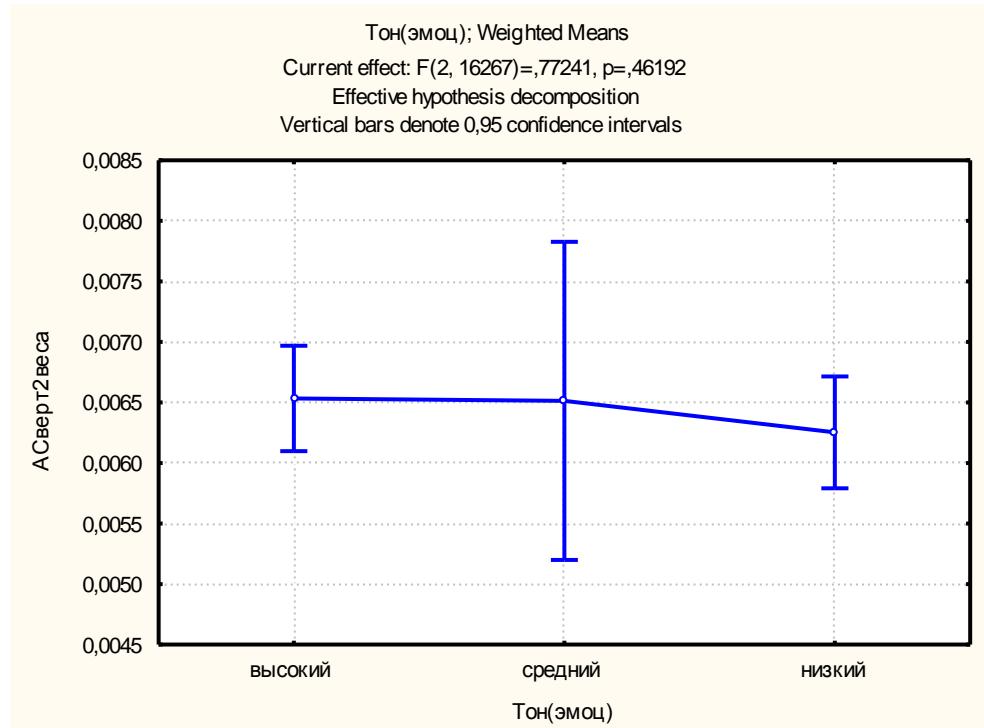


Рисунок 2.9 – Влияние фактора "Эмоциональная тональность" на обобщенный показатель полезности комментариев к законопроекту "О полиции"

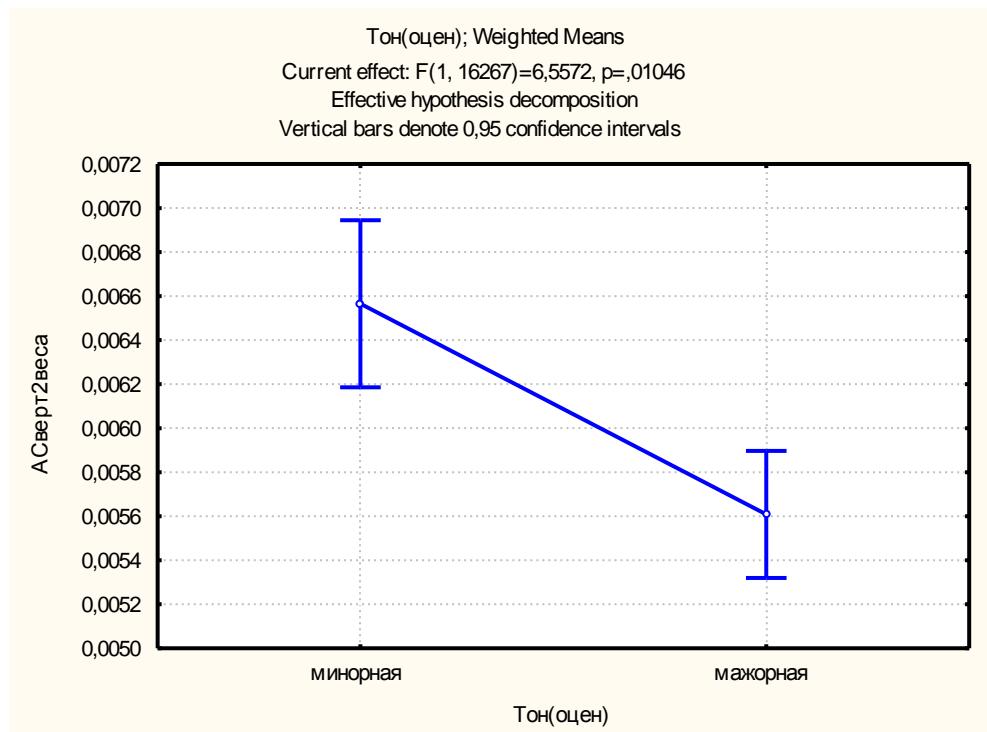


Рисунок 2.10 – Влияние фактора "Оценочная тональность" на обобщенный показатель полезности комментариев к законопроекту "О полиции"

Дисперсионный анализ на примере комментариев к законопроекту "О полиции" показал, что значимое влияние на обобщенный показатель полезности оказывают факторы "Статья", "Стиль" и "Оценочная тональность", а фактор "Эмоциональная тональность" незначим.

При переходе от одного обсуждаемого законопроекта к другому состав значимых атрибутов может изменяться. Так, для законопроекта "Об образовании" незначимы уже два фактора: "Эмоциональная тональность" и "Оценочная тональность" (таблица 2.15).

Таблица 2.15 – Результаты дисперсионного анализа для комментариев к законопроекту "Об образовании"

Эффект	SS	Степ. свободы	MS	F	p
Статья	0,684	139	0,005	5,64	0,000*
Стиль	0,143	4	0,036	40,97	0,000*
Тон(эмоц)	0,002	2	0,001	1,1	0,333
Тон(оцен)	0,001	1	0,001	1,62	0,204

Шаги 6, 9 и проверка условия 10 обеспечивают итерацию шагов 7 и 8 по всем классам, сформированным комбинацией значимых атрибутов $\{a_s\}$.

Шаги 7 и 8 необходимы для получения аналитических выражений функций предельной полезности комментариев $\{com_{v,j_v}\}$.

На шаге 7 в классах COM_v , сформированных комбинацией выявленных на шаге 5 значимых атрибутов $\{a_s\}$, производится ранжирование комментариев $\{com_{v,j_v}\}$ по убыванию обобщенного показателя полезности u_{v,j_v} , полученного для каждого комментария на шаге 2.

На шаге 8 полученные кривые аппроксимируются логарифмической функцией $g_v(Q_v)$. Пример аппроксимации, выполненной в пакете Excel 2007, представлен на рисунках 2.11 и 2.12 [102, 103].

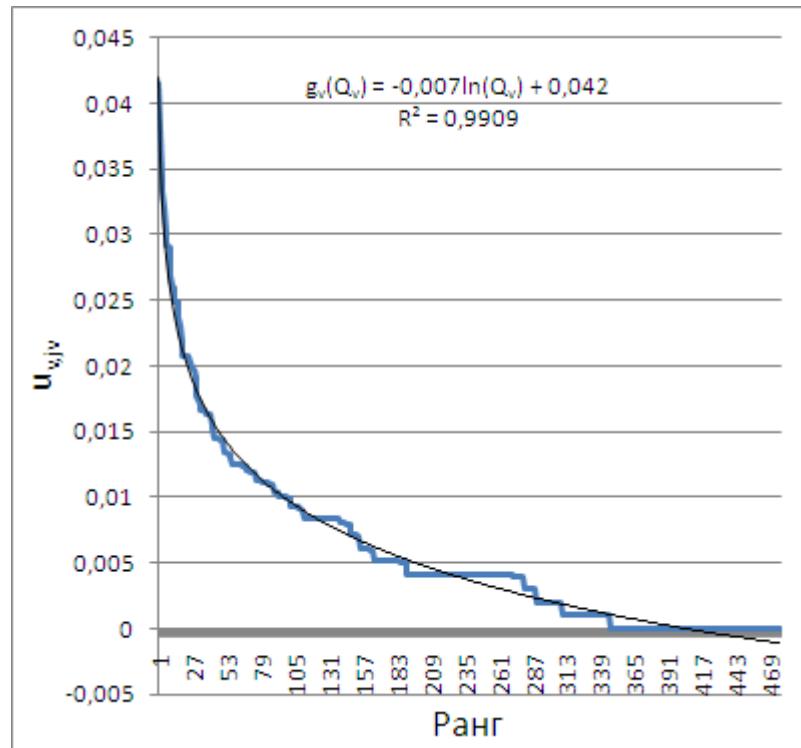


Рисунок 2.11 – Пример аппроксимации графика предельной полезности комментариев класса "2-Науч-Минор" к законопроекту "О полиции" логарифмической кривой с параметрами и достоверностью аппроксимации

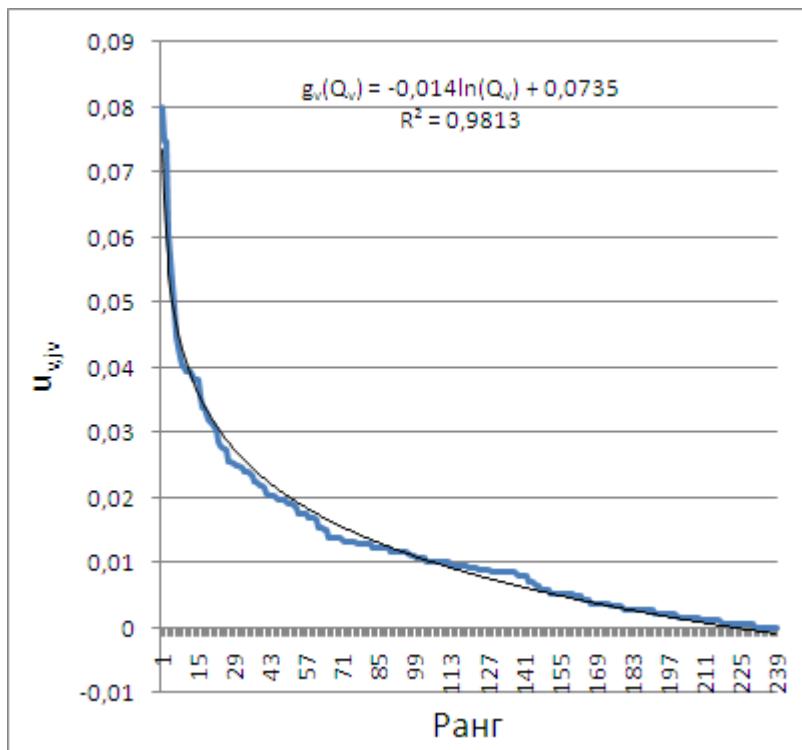


Рисунок 2.12 – Пример аппроксимации графика предельной полезности комментариев класса "49-Делов" к законопроекту "Об образовании" логарифмической кривой с параметрами и достоверностью аппроксимации

Результатом шага 8 является совокупность параметров всех функций предельной полезности $g_v(Q_v)$ (таблицы 2.9 и 2.10 – столбцы "A_v", "B_v", "Q_{v_max}", "Достоверность аппроксимации").

На шаге 11 определяются ограничения, характеризующие массив комментариев:

- максимальное число комментариев Q_{v_max} в классах, сформированных комбинацией значимых атрибутов $\{a_s\}$ (таблицы 2.16, 2.17: столбец "Q_{v_max}");
- средняя длина комментария в классе L_v (слов) (таблицы 2.16, 2.17: столбец "L_v").

Также на указанном шаге задаются стоимостные параметры:

- среднее время чтения одного слова $T1$ (на основе темпа чтения комментариев);
- бюджет времени Td с учетом ограничения $0 \leq Td \leq Td_{max}$.

Таблица 2.16 – Параметры функций предельной полезности и ограничения для классов комментариев COM_v к законопроекту "О полиции" (фрагмент)

v	Статья	Стиль	Тон.(оцен.)	A_v	$B_v (-B_v \cdot \ln Q)$	Достов. аппр.	Q_{v_max}	L_v
...
132	18	газет.-публ.	минорная	0,0303	0,009	0,9234	28	71,32
133	19	газет.-публ.	минорная	0,0176	0,005	0,9626	36	49,94
134	20	газет.-публ.	минорная	0,0221	0,006	0,9284	25	49,08
135	21	газет.-публ.	минорная	0,023	0,008	0,9344	20	60,7
136	22	газет.-публ.	минорная	0,0262	0,008	0,9526	30	58,73
137	23	газет.-публ.	минорная	0,051	0,012	0,5452	97	57,36
138	24	газет.-публ.	минорная	0,0381	0,011	0,8649	45	55,38
139	25	газет.-публ.	минорная	0,0257	0,007	0,9164	43	62,53
140	26	газет.-публ.	минорная	0,043	0,011	0,6124	72	58,28
141	27	газет.-публ.	минорная	0,0262	0,009	0,9342	26	56,85
...

Таблица 2.17 – Параметры функций предельной полезности и ограничения для классов комментариев COM_v к законопроекту "Об образовании" (фрагмент)

v	Статья	Стиль	A_v	$B_v (-B_v \cdot \ln Q)$	Достов. аппр.	Q_{v_max}	L_v
...
180	43	научный	0,0483	0,011	0,9737	59	34,32
181	44	научный	0,0562	0,013	0,9488	106	70,42
182	45	научный	0,2312	0,083	0,9489	20	40,9
183	46	научный	0,409	0,126	0,547	34	76,94
184	47	научный	0,0533	0,011	0,9643	70	83,91
185	48	научный	0,0419	0,011	0,9243	51	40,22
186	49	научный	0,0849	0,016	0,9301	221	58
187	50	научный	0,0353	0,01	0,9746	29	62,45
188	51	научный	0,0279	0,008	0,9776	48	47,65
189	52	научный	0,0453	0,015	0,7755	27	47
...

Совокупность параметров, полученных в результате выполнения шагов 8 и 11, используется на шаге 12 для расчета с использованием метода множителей Лагранжа квот Q_v для классов комментариев COM_v .

Основой для расчета выступает система (2.27), составляемая на основе параметров функций предельной полезности комментариев, стоимостных параметров и ограничений для конкретного законопроекта. Фрагмент системы

уравнений при $Td = 0,5 \cdot Td_{\max}$ с неизвестными параметрами выборки полезных комментариев к законопроекту "О полиции" представлен выражением:

$$\left\{ \begin{array}{l} 0,5 \cdot 348800 = \sum_{v=1}^{472} L_v \cdot 0,3 \cdot Q_v; \\ \frac{0,0756 - 0,013 \cdot \ln(Q_1)}{97,77 \cdot 0,3} = \frac{0,042 - 0,007 \cdot \ln(Q_2)}{73,45 \cdot 0,3}; \\ \frac{0,042 - 0,007 \cdot \ln(Q_2)}{73,45 \cdot 0,3} = \frac{0,0508 - 0,009 \cdot \ln(Q_3)}{84,82 \cdot 0,3}; \\ \dots \\ \frac{0,0132 - 0,005 \cdot \ln(Q_{236})}{58,27 \cdot 0,3} = \frac{0,0128 - 0,004 \cdot \ln(Q_{237})}{70,29 \cdot 0,3}; \\ \dots \\ \frac{0,003 - 0,004 \cdot \ln(Q_{471})}{10 \cdot 0,3} = \frac{0,001 - 0,001 \cdot \ln(Q_{472})}{15 \cdot 0,3}; \\ 1 \leq Q_1 \leq 496; \\ 1 \leq Q_2 \leq 476; \\ \dots \\ 1 \leq Q_{237} \leq 14; \\ \dots \\ 1 \leq Q_{471} \leq 1; \\ 1 \leq Q_{472} \leq 1. \end{array} \right. \quad (2.41)$$

При составлении системы (2.41) для нумерации классов комментариев был задан порядок – по убыванию мощности классов $|COM_v|$. Результатом решения системы (2.41), выполненного в пакете Mathcad 14, является вектор квот для классов комментариев:

$$Q = (248, 238, \dots, 7, \dots, 1). \quad (2.42)$$

На шаге 13 полученные на шаге 12 квоты применяются к полученным на шаге 7 ранжировкам, вошедшие в квоту комментарии выбираемого аналитиком класса составляют выборку полезных комментариев.

Ознакомление с выборкой полезных комментариев класса COM_v производится на шаге 14 в пределах квоты Q_v в порядке убывания обобщенного показателя полезности u_{v,j_v} . Например, для класса COM_{237} (статья="45", стиль="газетно-публицистический", оценочная тональность="минорная") при квоте $Q_{237} = 7$ для $Td = 0,5 \cdot Td_{\max}$ ознакомление с комментариями в соответствии с таблицей 2.18 следует осуществлять следующем порядке:

$$16008 \rightarrow 316 \rightarrow 5051 \rightarrow 17204 \rightarrow 7059 \rightarrow 7243 \rightarrow 13143. \quad (2.43)$$

Таблица 2.18 – Комментарии класса "45-Газет-Минор", отранжированные в порядке убывания обобщенного показателя полезности

Ранг	Идентификатор комментария	Обобщенный показатель полезности
1	16008	0,010305
2	316	0,00931
3	5051	0,009131
4	17204	0,008316
5	7059	0,008136
6	7243	0,005968
7	13143	0,005152
8	17717	0,005152
9	10404	0,002984
10	7061	0,001989
11	1689	0,000995
12	15939	0,000995
13	16492	0,000995
14	1816	0

Предложенный алгоритм формирования выборки полезных комментариев позволяет:

– находить такое множество комментариев $UCS = \{com_{v,j_v}\}$,

$UCS \subseteq COM$, представляющее собой выборку полезных комментариев, для которого выполняется система требований:

$$\begin{cases} |UCS| \rightarrow \min; \\ \sum_{v=1}^w \sum_{j_v=1}^{h_v} u_{v,j_v} \rightarrow \max; \\ Td = const, \end{cases} \quad (2.44)$$

– учесть неравномерность распределения высказываний в комментариях и предпочтения интернет-пользователей при формировании выборки полезных комментариев на основе формирования и применения в расчетах свертки:

$$u_{v,j_v} = \alpha_p \cdot \frac{p_{v,j_v}}{p_{\max}} + \alpha_l \cdot \frac{l_{v,j_v}}{l_{\max}}, \quad (2.45)$$

которая составляется с учетом распределений данных параметров и представляет собой обобщенный показатель полезности комментариев;

– определять квоты комментариев в классах на основе расчета с применением метода множителей Лагранжа, исходными данными для которого являются: аппроксимированные логарифмические функции предельной полезности комментариев в классах; стоимостные параметры, характеризующие общий бюджет времени и темп изучения комментариев; ограничения, определяющие среднюю длину и максимальное количество комментариев в классах;

– обеспечить получение данных для проведения качественной детализации на основе формирования выборки полезных комментариев в ходе оперативной аналитической обработки комментариев.

Выводы по 2 разделу

1. Предложена модель оперативной аналитической обработки комментариев с операциями детализации на основе выявления типичных и полезных текстов, позволяющая:

- осуществлять анализ массива комментариев как массива многомерных данных с использованием стандартных OLAP-операций и устанавливать порядок обработки классов комментариев с учетом интенсивности мнений;
- производить операцию погружения в ходе анализа массива комментариев с использованием качественной детализации на основе выборки типичных комментариев, учитывающей содержательное подобие текстов комментариев, сходство со сводным рефератом кластера и выделенное директивное время на обработку комментариев для их приоритетного отбора;
- осуществлять операцию погружения в ходе анализа массива комментариев с использованием качественной детализации на основе выборки полезных комментариев, учитывающей неравномерное распределение конструктивных высказываний в комментариях, уровень поддержки интернет-пользователей и выделенное директивное время на обработку комментариев для их приоритетного отбора.

2. Предложен алгоритм формирования выборки типичных комментариев, позволяющий:

- находить такое множество комментариев $TCS = \{com_{v, j_v}\}$, $TCS \subseteq COM$, представляющее собой выборку типичных комментариев, для которого выполняется система требований (2.32);
- учесть содержательное подобие комментариев и сходство с типичным представителем мнения при формировании выборки типичных комментариев на основе кластеризации с использованием матриц подобия с последующим определением в кластерах комментариев, обладающих наилучшим сходством со сводным рефератом кластера;

- аккумулировать в комментариях выборки типичные высказывания в наиболее удаленных по содержанию кластерах подобных комментариев;
- обеспечить получение данных для проведения качественной детализации на основе формирования выборки типичных комментариев в ходе оперативной аналитической обработки комментариев.

3. Предложен алгоритм формирования выборки полезных комментариев, позволяющий:

- находить такое множество комментариев $UCS = \{com_{v, j_v}\}$, $UCS \subseteq COM$, представляющее собой выборку полезных комментариев, для которого выполняется система требований (2.44);
- учесть неравномерность распределения высказываний в комментариях и предпочтения интернет-пользователей при формировании выборки полезных комментариев на основе формирования и применения в расчетах свертки (2.45), которая составляется с учетом распределений данных параметров и представляет собой обобщенный показатель полезности комментариев;
- определять квоты комментариев в классах на основе расчета с применением метода множителей Лагранжа, исходными данными для которого являются: аппроксимированные логарифмические функции предельной полезности комментариев в классах; стоимостные параметры, характеризующие общий бюджет времени и темп изучения комментариев; ограничения, определяющие среднюю длину и максимальное количество комментариев в классах;
- обеспечить получение данных для проведения качественной детализации на основе формирования выборки полезных комментариев в ходе оперативной аналитической обработки комментариев.

3 ОЦЕНИВАНИЕ АДЕКВАТНОСТИ МОДЕЛИ И ЭФФЕКТИВНОСТИ РАЗРАБОТАННЫХ АЛГОРИТМОВ

Оценка полученных результатов осуществлялась путем определения адекватности модели и эффективности алгоритмов.

Оценивание адекватности модели проводилось посредством сравнения типичных высказываний, выделенных экспертами в ходе подготовки аналитического обзора, с высказываниями, полученными с использованием модели.

При этом учитывались теоретические положения о том, что в тексте, как в сложной системе, отсутствуют математические свойства. Текст не обладает аддитивностью, коммутативностью и ассоциативностью. Текст обладает субъективной ценностью и интерпретируемостью – каждый получатель информации воспринимает содержание текста по-своему [64, 104].

Применительно к оценке модели эти теоретические положения ограничивают возможность точного вычисления адекватности модели. Тем не менее для оценки адекватности использовалось выражение:

$$r_v = \frac{q_v - k_v}{k_v}, \quad (3.1)$$

где k_v – мощность класса комментариев ω_v , q_v – мощность подмножества класса комментариев ω_v (выборки заданного типа, подмножества исходного массива), обеспечивающая охват всех типичных высказываний в классе. Указанное выражение позволяет оценить адекватность модели на основе оценки обобщающей способности формируемых в соответствии с разработанной моделью выборок исследуемых классов.

Оценивание эффективности разработанных алгоритмов проводилось по следующим направлениям:

- оценивание среднего темпа прироста полноты отбора высказываний в ходе обработки комментариев с применением выборок и сравнение его с темпом прироста полноты при исходной структуре классов массива;
- оценивание свойств алгоритмов.

Для оценивания адекватности модели и эффективности алгоритмов были исследованы статьи законопроектов "О полиции" и "Об образовании" [94, 105, 106]. В ходе исследования было установлено, что статьи законопроектов содержат различное количество пунктов и подпунктов, что обуславливает большее либо меньшее разнообразие высказываний (приложение А, таблицы А.1 и А.2).

При этом распределение статей по количеству объектов комментирования подчинено экспоненциальному закону (приложение А, таблицы А.3 и А.4, рисунки А.1 и А.2).

Для экспериментов были отобраны классы комментариев к статьям, содержащим наибольшее, наименьшее и среднее количество объектов комментирования, т.е. по три класса для двух законопроектов. В данных классах был произведен экспертный анализ комментариев, выделены содержащиеся в них высказывания (приложение Б, таблицы Б.1–Б.6) и составлены распределения высказываний по комментариям классов (приложение Б, таблицы Б.7–Б.12).

3.1 Оценивание и сравнение обобщающих и накапливающих способностей выборок типичных и полезных комментариев

Исследование обобщающих и накапливающих способностей выборок проводилось путем сравнения высказываний, полученных с использованием алгоритмов, и типичных высказываний, включенных экспертами в аналитический обзор к законопроекту "О полиции" (таблица 3.1).

Таблица 3.1 – Типичные высказывания, выделенные экспертами и включенные в аналитический обзор

v	m_v	Содержание высказывания $sent_{v,m_v}$
3-12	23	Заменить слова "содействие организаторам спортивных, зрелищных мероприятий..." на "охрана общественного порядка"
	28	Убрать обязанности по контролю соблюдения миграционного законодательства
	29	Исключить право проведения экспертиз по уголовным и административным делам
	30	Исключить охрану на договорной основе имущества граждан и организаций
7-42	2	Уточнить основания для увольнения в связи с организационно-штатными мероприятиями
	3	Убрать право выбора основания увольнения
6-34	2	Конкретизировать требования (статус, права и обязанности, уровень образования, степень подготовки, экипировка и пр.) к привлекаемым к охране общественного порядка и пр. задачам иных сотрудников органов внутренних дел (ОВД) и стажеров
	5	Предоставить гражданину право удостовериться, что привлеченный сотрудник уполномочен совершать в отношении него действия

Под обобщающей способностью выборки понимается зависимость количества типичных высказываний, вошедших в выборку, от ее мощности (выражение (3.1)).

Для оценивания обобщающих способностей выборок комментариев для каждого из типичных высказываний были проанализированы мощности подмножества класса массива комментариев и выборки каждого типа, обеспечивающие охват данных высказываний. В качестве характеристики обобщающей способности выборок в соответствии с выражением (3.1) было взято максимальное из полученных значений для класса, выраженное в процентах (таблица 3.2).

На основе полученных данных было произведено сравнение обобщающих способностей выборок для каждого из исследуемых классов. Ступенчатый характер графиков для отдельных классов обусловлен пошаговым изучением комментариев, при этом переход на более высокий уровень происходит, если комментарий содержит одно или несколько типичных высказываний (рисунки 3.1–3.3).

Таблица 3.2 – Оценивание обобщающих способностей выборок на примере классов законопроекта "О полиции"

		Класс, v		Типичное высказывание, включенное экспертами в аналитический обзор, m_v		Мощность подмножества класса массива комментариев, обеспечивающая охват типичных высказываний		Доля от мощности класса массива комментариев, обеспечивающая охват типичных высказываний, %		Доля от мощности класса массива комментариев, обеспечивающая охват всех типичных высказываний, %		Мощность выборки типичных комментариев, обеспечивающая охват типичных высказываний		Доля от max мощности выборки типичных комментариев, обеспечивающая охват типичных высказываний, %		Доля от max мощности выборки типичных комментариев, обеспечивающая охват всех типичных высказываний, %		Мощность выборки полезных комментариев, обеспечивающая охват типичных высказываний		Доля от max мощности выборки полезных комментариев, обеспечивающая охват типичных высказываний, %		Доля от max мощности выборки полезных комментариев, обеспечивающая охват всех типичных высказываний, %						
3-12	23	20	34,5	50	11	19	19	14	24,1	24,1	3,6	1	3,6	3,6	1	3,6	3,6	1	3,6	33,3	1	11,1	11,1	3	33,3			
	28	27	46,6		8	13,8		9	15,5		3,6	1	3,6		1	3,6		1	3,6		1	5,2		1	1,7			
	29	28	48,3		9	15,5		1	1,7		44,4	1	11,1		1	11,1		1	11,1		44,4	1	11,1	3	33,3			
	30	29	50		1	1,7		3	5,2																			
7-42	2	1	3,6	3,6	1	3,6	3,6	1	3,6	3,6				3,6	1	3,6	3,6	1	3,6	33,3			33,3			33,3		
	3	1	3,6		1	3,6		1	3,6						1	3,6		1	3,6									
6-34	2	1	11,1	44,4	1	11,1	11,1	1	11,1	11,1				11,1	1	11,1	11,1	1	11,1	11,1			11,1			11,1		
	5	4	44,4		1	11,1		3	33,3																			

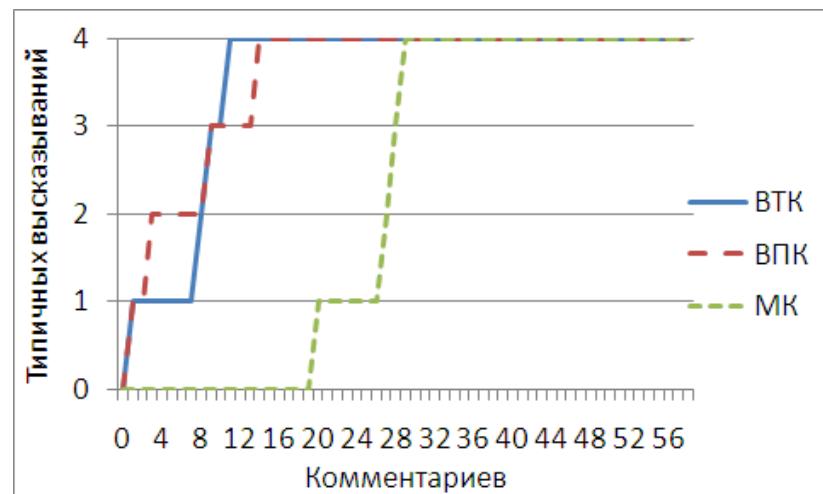


Рисунок 3.1 – Сравнение обобщающих способностей выборок типичных и полезных комментариев для класса "3-12" к законопроекту "О полиции"

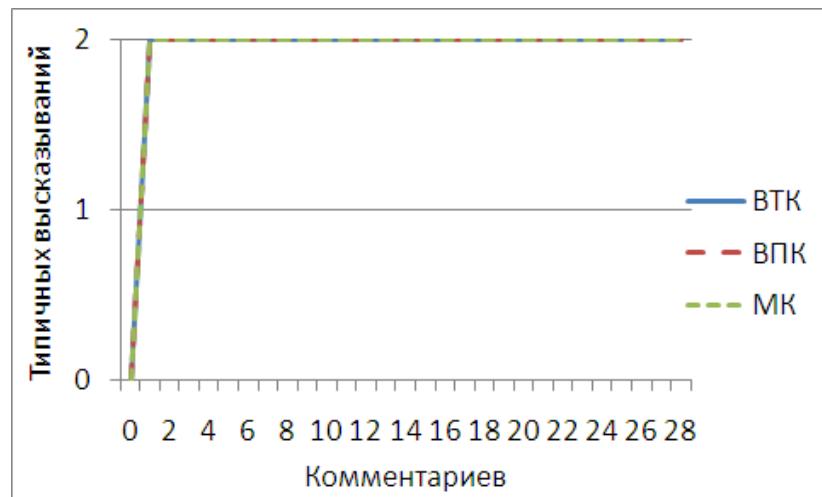


Рисунок 3.2 – Сравнение обобщающих способностей выборок типичных и полезных комментариев для класса "7-42" к законопроекту "О полиции"

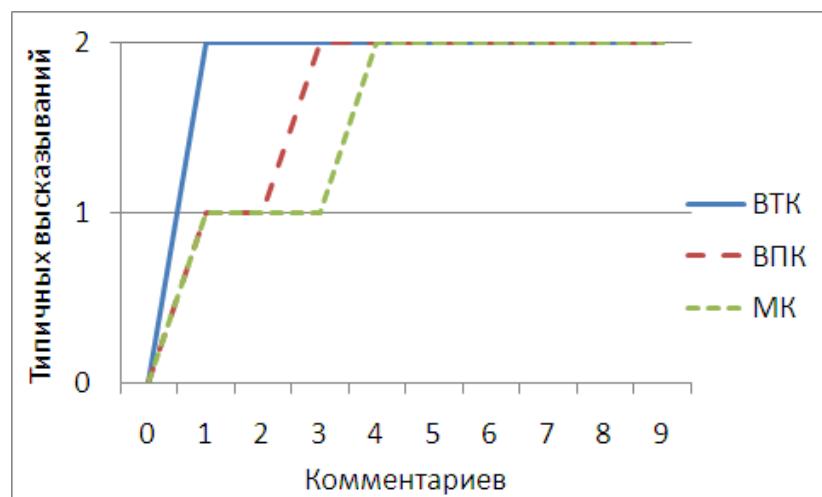


Рисунок 3.3 – Сравнение обобщающих способностей выборок типичных и полезных комментариев для класса "6-34" к законопроекту "О полиции"

Из графиков видно, что:

- модель обеспечивает выделение высказываний, совпадающих по смыслу с высказываниями, приведенными в обзоре в качестве типичных;
- процедура ознакомления эксперта с комментариями на основе классификации и выборок обеспечивает более раннее знакомство с типичными высказываниями и позволяет понять суть наиболее общих предложений при небольшом бюджете времени на обработку;

в) выделение типичных высказываний с использованием модели осуществляется в среднем на 20 % быстрее по сравнению с выделением этих же высказываний путем последовательного чтения.

При этом по результатам сравнения двух типов выборок установлено, что выборка типичных комментариев позволяет ознакомиться с типичными высказываниями в изучаемом классе на меньшем в среднем на 10 % от мощности класса объеме выборки (рисунок 3.4), то есть обладает большей обобщающей способностью.

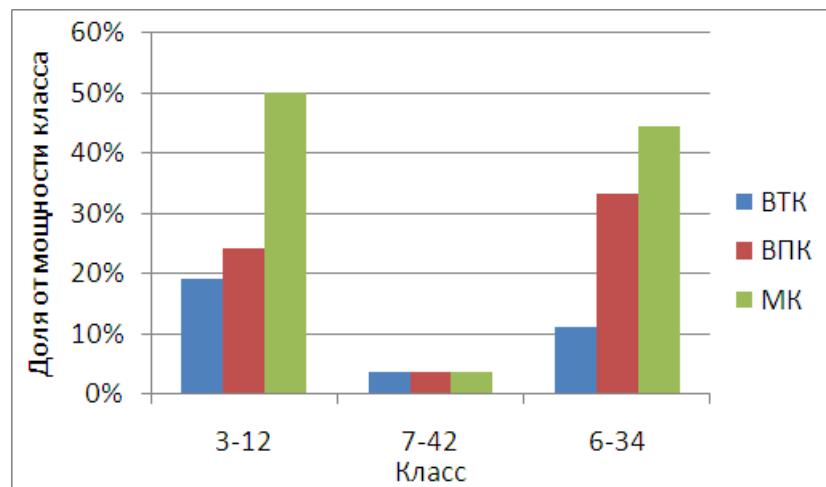


Рисунок 3.4 – Мощности выборок, обеспечивающие охват типичных комментариев для классов законопроекта "О полиции"

Под накапливающей способностью выборки понимается зависимость количества высказываний, вошедших в выборку, от ее мощности. Для оценивания накапливающих способностей выборок был проведен анализ данной зависимости для исследуемых классов (рисунки 3.5–3.7).

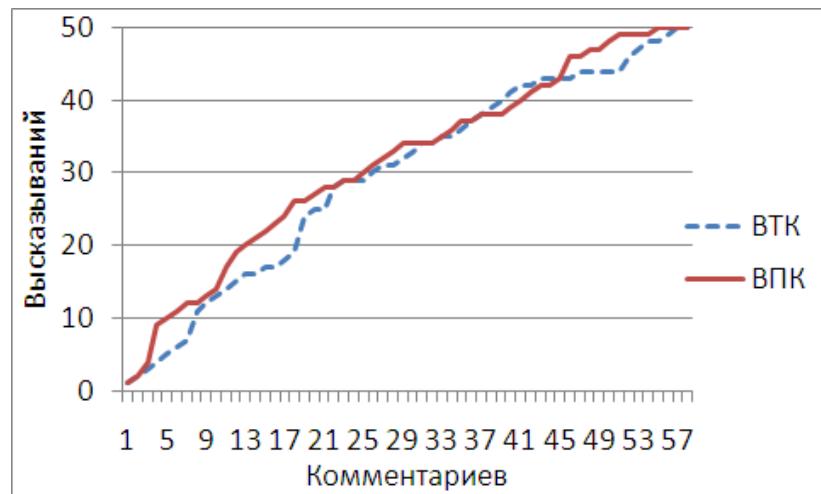


Рисунок 3.5 – Сравнение накапливающих способностей выборок типичных и полезных комментариев для класса "3-12" к законопроекту "О полиции"

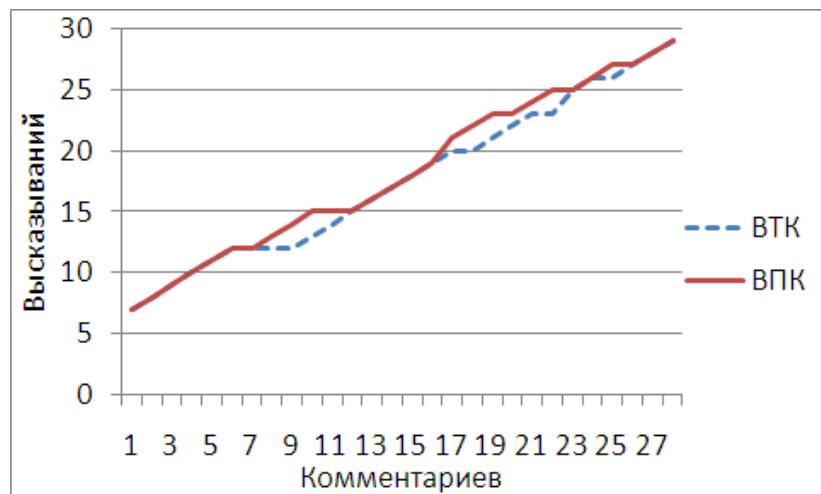


Рисунок 3.6 – Сравнение накапливающих способностей выборок типичных и полезных комментариев для класса "7-42" к законопроекту "О полиции"

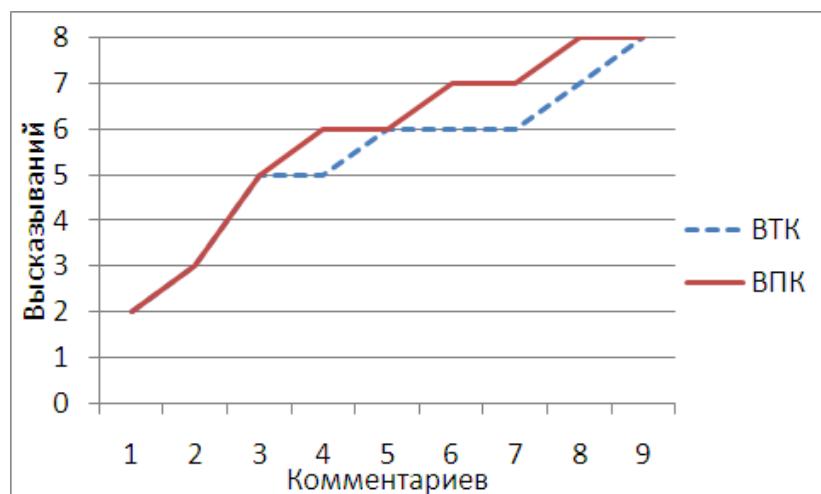


Рисунок 3.7 – Сравнение накапливающих способностей выборок типичных и полезных комментариев для класса "6-34" к законопроекту "О полиции"

Исследования показали, что при равных мощностях выборок выборка полезных комментариев содержит в среднем на 2–5 % больше высказываний, то есть обладает большей способностью по их накоплению.

При сравнении обобщающих и накапливающих способностей выборок было выявлено, что выборка типичных комментариев более эффективна для ознакомления с обобщенными высказываниями, представленными в комментариях изучаемого класса, а выборка полезных – для ознакомления с разнообразием высказываний в классе комментариев.

3.2 Оценивание эффективности алгоритма формирования выборки типичных комментариев

Оценивание алгоритмов проводилось по временной эффективности алгоритмов и по временной функции сложности.

Выбор временной эффективности для оценки алгоритмов связан с тем, что анализ предметной области показал отсутствие алгоритмов и программных комплексов, решающих подобные задачи. В связи с этим, временная эффективность алгоритмов определялась на основе сравнения темпов прироста полноты отбора высказываний в исходном, не отранжированном по типичности (полезности) множестве комментариев и выборках при одинаковом бюджете времени.

Для оценивания временной эффективности алгоритма формирования выборки типичных комментариев в исследуемых классах в соответствии с алгоритмом были определены множества комментариев, предоставляемые для изучения в соответствии с выделенным бюджетом времени с шагом $\Delta t = 0,05 \cdot Td_{\max}$.

Полученные множества пошагово сравнивались с множествами не отранжированных по типичности комментариев при одинаковом бюджете времени по четырем видам полноты (таблица 3.3).

Таблица 3.3 – Виды полноты отбора высказываний в комментариях, включенных в выборку

По учету веса высказываний	По идентификации повторов высказываний	
	Без идентификации повторов высказываний	С идентификацией повторов высказываний
Без учета веса высказываний	R1	R3
С учетом веса высказываний	R2	R4

Для проведения расчетов полноты отбора высказываний в исходном и отранжированном по типичности множествах комментариев в пределах квот, полученных с шагом $\Delta t = 0,05 \cdot Td_{\max}$, предварительно в выбранных для экспериментов классах исследовалось пошаговое (при добавлении по одному комментарию) накопление полноты. Примеры расчетов для классов комментариев "7-42-Н-В-Ми" к законопроекту "О полиции" и "1-11-ОД-Н-Ми" к законопроекту "Об образовании" приведены в таблицах В.1–В.4 приложения В.

На основе данных, полученных в ходе предварительных расчетов пошагового накопления полноты и квот для классов комментариев, исследовался средний темп прироста полноты каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в выбранных для проведения экспериментов классах (приложение В, таблицы В.5–В.10).

В ходе экспериментов было установлено, что средний прирост полноты зависит от способа ее расчета, в зависимости от учета веса высказываний и идентификации повторов высказываний. Пример зависимости полноты отбора высказываний разных типов от выделенного бюджета времени представлен на рисунке 3.8.

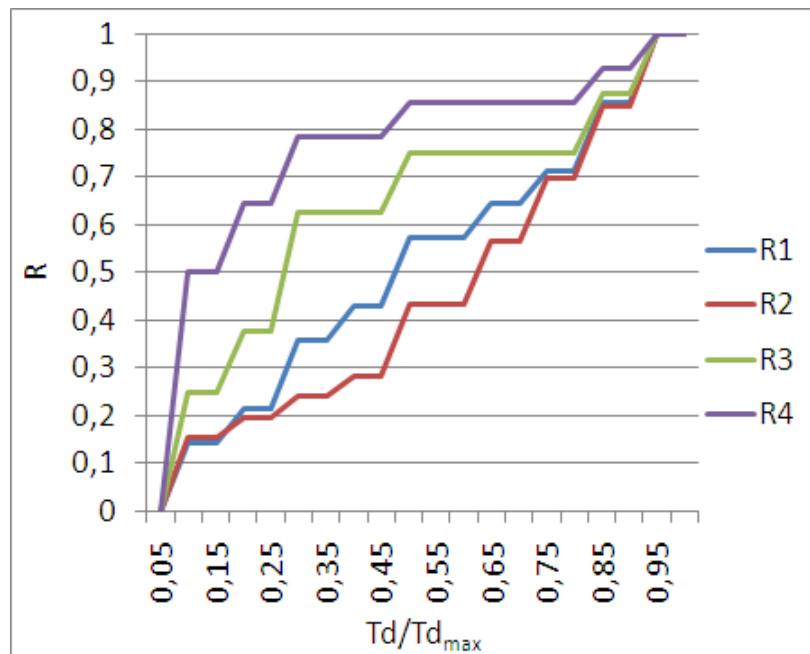


Рисунок 3.8 – Зависимость полноты разного типа от выделенного бюджета времени на изучение комментариев для класса "6-34-Н-Н-Ми" к законопроекту "О полиции"

Эксперименты показали, что накопление полноты отбора высказываний во включенных в соответствии с алгоритмом в выборку комментариях в начальную фазу их обработки (до половины бюджета времени, в зависимости от класса) происходит быстрее (рисунок 3.9).

При этом на данной фазе полнота отбора высказываний в относительных величинах может значительно (в 1,5 и более раз) превосходить полноту отбора в неотранжированных по типичности комментариях исходного массива, что объясняется наличием типичных (присутствующих в большом числе комментариев) высказываний в комментариях с высокими рангами (рисунок 3.10).

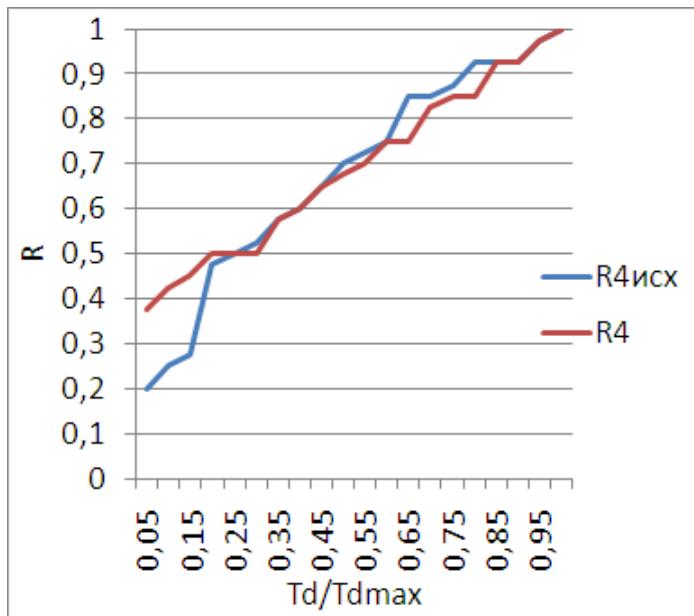


Рисунок 3.9 – Накопление полноты отбора высказываний в комментариях выборки R4 с ростом бюджета времени в сравнении с полнотой отбора в неотранжированных по типичности комментариях R4_{исх} исходного массива на примере класса "7-42-Н-В-Ми" к законопроекту "О полиции"

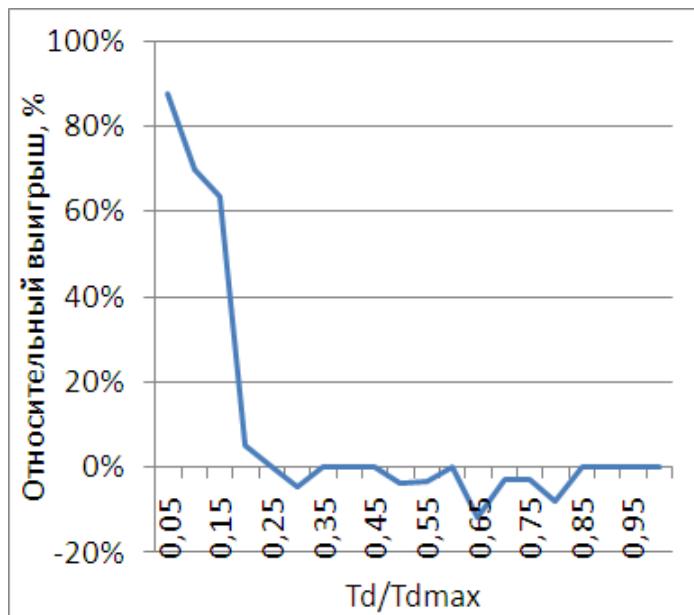


Рисунок 3.10 – Относительный выигрыш в накоплении R4 по сравнению с R4_{исх} на примере класса "7-42-Н-В-Ми" к законопроекту "О полиции"

Оцениваемый алгоритм формирования выборки типичных комментариев показал наилучшие результаты при оценке по среднему приросту полноты вида R4 при $\Delta t = 0,05 \cdot Td_{max}$ (рисунки 3.11–3.12).

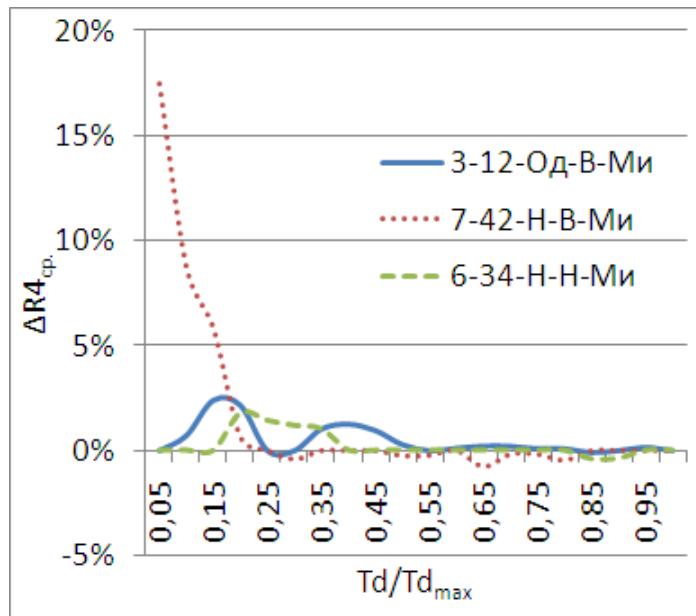


Рисунок 3.11 – Абсолютный выигрыш в среднем приросте полноты вида R4 при $\Delta t = 0,05 \cdot Td_{max}$ для выборки типичных комментариев к трем классам законопроекта "О полиции" с различным числом объектов комментирования

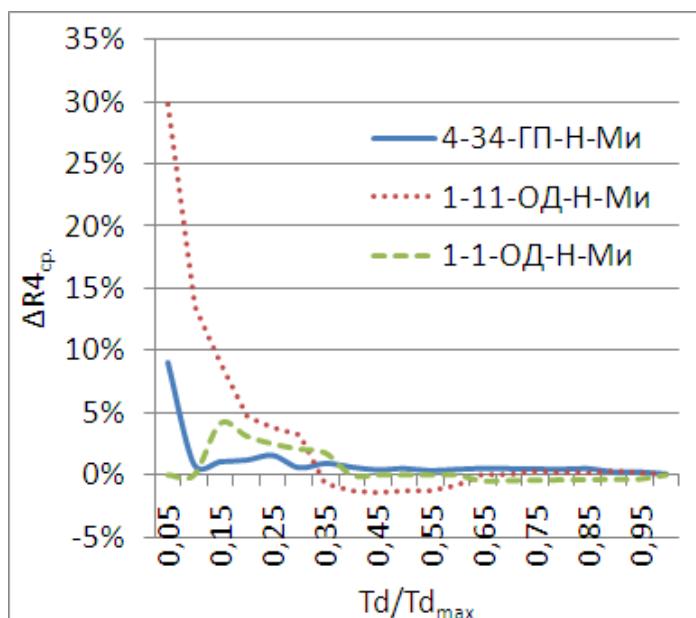


Рисунок 3.12 – Абсолютный выигрыш в среднем приросте полноты вида R4 при $\Delta t = 0,05 \cdot Td_{max}$ для выборки типичных комментариев к трем классам законопроекта "Об образовании" с различным числом объектов комментирования

Эксперименты показали, что для различных законопроектов и статей с различным количеством объектов комментирования разработанный алгоритм позволяет обеспечить опережающий темп изучения высказываний экспертом, как правило, лишь в первую половину от максимального бюджета времени, что связано с тем, что значительная часть комментариев содержит по одному уникальному высказыванию.

В результате оценивания эффективности алгоритма формирования выборки типичных комментариев можно сделать следующие выводы:

- высказывания распределены по комментариям неравномерно, в проанализированных классах встречаются высказывания, присутствующие в нескольких комментариях, однако доля таких высказываний невелика, значительное число высказываний уникально и содержится только в одном комментарии;
- использование меры подобия в качестве прототипа семантической метрики и сводного реферата в качестве типичного представителя мнения позволяет на основе оценки содержательного подобия комментариев группировать их в кластеры, выявлять наличие повторяющихся высказываний и обеспечить путем оценки сходства с типичным представителем мнения в кластере приоритетную обработку типичных комментариев, содержащих часто встречающиеся высказывания;
- накопление полноты отбора высказываний во включенных в соответствии с алгоритмом в выборку комментариях в начальную фазу их обработки (до половины бюджета времени, в зависимости от класса) происходит быстрее, при этом на данной фазе полнота отбора высказываний в относительных величинах может значительно (в 1,5 и более раз) превосходить полноту отбора в неотранжированных по типичности комментариях исходного массива, что объясняется наличием типичных (присутствующих в большом числе комментариев) высказываний в комментариях с высокими рангами;

– направленный отбор комментариев с учетом содержательного подобия и сходства с типичным представителем мнения, осуществляемый с использованием разработанного алгоритма, позволяет обеспечить более высокий средний темп прироста полноты отбора высказываний за счет приоритетной обработки комментариев, содержащих типичные высказывания.

3.3 Оценивание эффективности алгоритма формирования выборки полезных комментариев

Оценивание алгоритма также проводилось по временной эффективности и по временной функции сложности.

Для оценивания временной эффективности алгоритма формирования выборки полезных комментариев в исследуемых классах в соответствии с алгоритмом были определены множества комментариев, предоставляемые для изучения в соответствии с выделенным бюджетом времени с шагом $\Delta t = 0,05 \cdot Td_{\max}$.

В отличие от алгоритма формирования выборки типичных комментариев, формирующего квоты для одного класса, исследуемый алгоритм формирует квоты для требуемой комбинации классов.

Фрагмент таблицы с результатами расчета квот для законопроекта "О полиции", выполненного в пакете Mathcad 14, приведен в таблице 3.4. Семейство кривых, описывающих рост квот Q_v в классах с увеличением бюджета времени, представлено на рисунке 3.13.

Полученные множества пошагово сравнивались с множествами не отранжированных по полезности комментариев при одинаковом бюджете времени по четырем видам полноты (таблица 3.3).

Таблица 3.4 – Результаты расчета квот для комментариев к законопроекту "О полиции" в соответствии с выделенным бюджетом времени с шагом $\Delta t = 0,05 \cdot Td_{\max}$ (фрагмент, показаны первые по мощности 25 классов)

v	Статья	Стиль	Тон(оценка)	Q _{max}	Td / Td _{max}																				
					0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1	
1	13	науч.	минорная	496	2	12	42	60	94	118	140	166	190	248	234	261	285	309	334	393	410	406	435	496	
2	2	науч.	минорная	476	1	4	13	29	57	90	118	154	186	238	244	281	313	345	380	476	476	475	476	476	
3	1	науч.	минорная	385	1	4	11	21	36	66	92	114	137	192	180	207	230	253	278	340	363	348	377	385	
4	12	науч.	минорная	348	1	3	9	9	22	35	52	54	73	174	122	137	158	179	201	229	262	267	295	348	
5	2	газет.	минорная	317	4	9	29	16	45	82	95	121	138	158	168	186	202	218	237	261	287	285	308	317	
6	15	науч.	минорная	274	1	1	5	2	12	36	57	87	111	137	158	185	209	234	261	274	274	274	274	274	
7	9	науч.	минорная	262	1	1	4	2	10	24	47	65	87	131	128	153	175	198	222	262	262	262	262	262	
8	1	делов.	минорная	260	1	4	8	6	18	34	53	65	81	130	111	128	144	160	176	212	228	225	247	260	
9	2	делов.	минорная	259	1	4	8	7	23	38	55	68	83	129	113	130	146	162	179	227	244	228	251	259	
10	13	газет.	минорная	254	1	6	13	11	40	58	68	91	106	127	133	150	164	179	197	254	254	240	254	254	
11	5	науч.	минорная	251	1	6	16	14	33	44	56	65	77	125	104	117	130	143	158	190	211	199	220	251	
12	13	делов.	минорная	235	20	44	67	60	73	83	89	97	102	117	110	115	119	123	128	134	141	142	152	235	
13	1	газет.	минорная	213	1	6	14	13	21	31	46	55	67	106	92	104	116	128	143	159	192	181	201	213	
14	12	делов.	минорная	203	1	2	6	6	10	16	31	39	54	101	86	102	117	133	150	157	193	200	203	203	
15	37	науч.	минорная	192	1	1	3	3	4	12	31	51	73	96	118	140	162	184	192	192	192	192	192	192	
16	9	газет.	минорная	163	7	16	28	29	30	50	61	75	84	82	100	106	113	120	130	128	135	152	163	163	
17	37	газет.	минорная	159	1	4	7	9	10	21	23	27	33	79	57	64	75	85	93	89	97	130	147	159	
18	14	науч.	минорная	152	1	5	9	14	18	35	43	58	70	76	93	103	114	125	139	152	152	152	152	152	
19	12	газет.	минорная	146	1	6	14	17	28	42	47	62	72	73	95	103	113	123	135	142	146	146	146	146	
20	23	науч.	минорная	144	3	19	45	33	60	71	73	87	94	72	120	116	123	130	143	144	144	144	144	144	
21	44	газет.	минорная	144	8	28	49	35	55	64	68	77	83	72	96	98	102	107	114	119	124	129	139	144	
22	15	газет.	минорная	142	1	2	6	3	10	13	14	19	29	71	55	65	78	90	102	117	133	141	142	142	
23	26	науч.	минорная	124	1	1	4	2	7	10	16	20	31	62	55	65	76	88	102	124	124	124	124	124	
24	4	науч.	минорная	123	1	3	7	5	11	14	22	24	30	61	48	54	62	70	78	91	95	104	119	123	
25	30	науч.	минорная	123	17	31	38	36	43	45	50	51	54	61	58	59	61	63	64	66	67	72	79	123	
...		

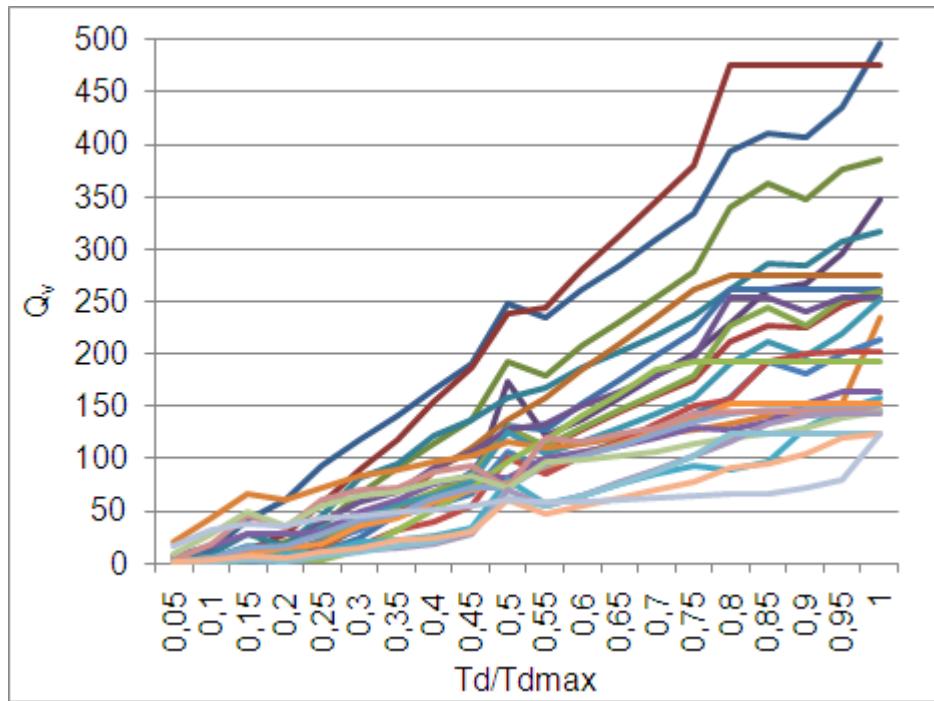


Рисунок 3.13 – Графики роста квот в классах с увеличением бюджета времени ($\Delta t = 0,05 \cdot Td_{max}$, показаны первые по мощности 25 классов, законопроект "О полиции")

Для проведения расчетов полноты отбора высказываний в исходном и отранжированном по полезности множествах комментариев в пределах квот, полученных с шагом $\Delta t = 0,05 \cdot Td_{max}$, предварительно в выбранных для экспериментов классах исследовалось пошаговое (при добавлении по одному комментарию) накопление полноты. Примеры расчетов для классов комментариев "7-42-Н-В-Ми" к законопроекту "О полиции" и "1-11-ОД-Н-Ми" к законопроекту "Об образовании" приведены в таблицах В.1, В.3 приложения В (для исходного множества, рассчитано ранее), Г.1, Г.2 приложения Г (для отранжированного по полезности множества).

На основе данных, полученных в ходе предварительных расчетов пошагового накопления полноты и квот для классов комментариев, исследовался средний темп прироста полноты каждого вида с шагом $\Delta t = 0,05 \cdot Td_{max}$ в выбранных для проведения экспериментов классах (приложение Г, таблицы Г.3–Г.8).

В ходе экспериментов было установлено, что средний прирост полноты зависит от способа ее расчета, в зависимости от учета веса высказываний и идентификации их повторов. Пример зависимости полноты отбора высказываний разных типов от выделенного бюджета времени представлен на рисунке 3.14.

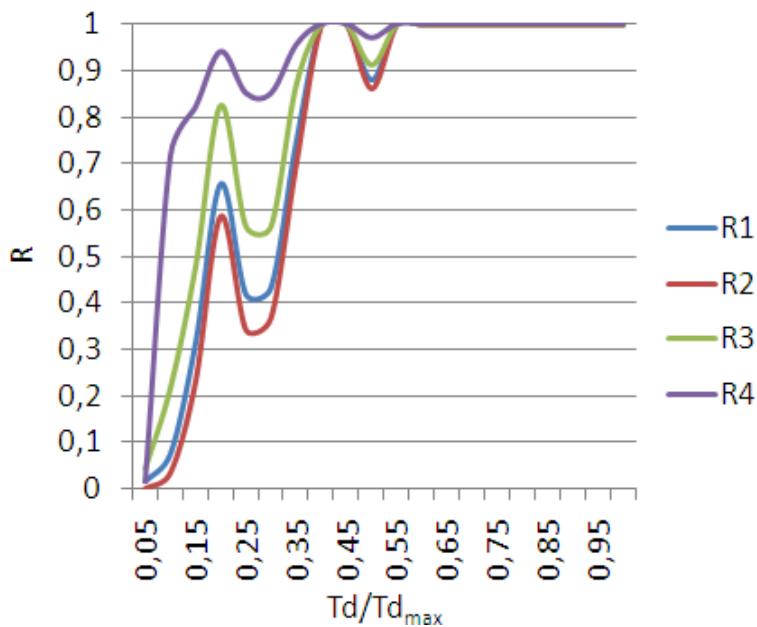


Рисунок 3.14 – Зависимость полноты разного типа от выделенного бюджета времени на изучение комментариев для класса "4-34-ГП-Н-Ми" к законопроекту "Об образовании"

Эксперименты показали, что накопление полноты отбора высказываний во включенных в соответствии с алгоритмом в выборку комментариях в начальную фазу их обработки (до половины бюджета времени, в зависимости от класса) происходит быстрее (рисунок 3.15).

При этом на данной фазе полнота отбора высказываний в относительных величинах может значительно (в 2 и более раз) превосходить полноту отбора в неотранжированных по полезности комментариях исходного массива, что объясняется наличием большого числа высказываний в комментариях с высокими рангами (рисунок 3.16).

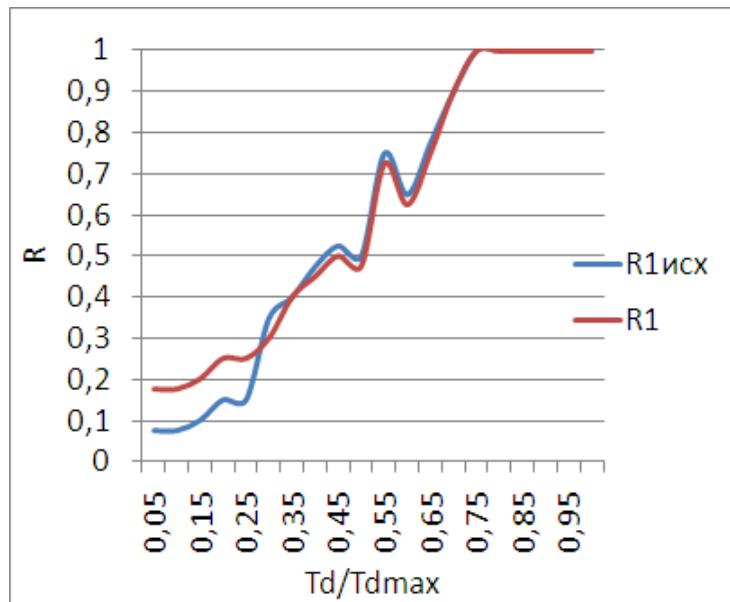


Рисунок 3.15 – Накопление полноты отбора высказываний в комментариях выборки R1 с ростом бюджета времени в сравнении с полнотой отбора в неотранжированных по полезности комментариях R1исх исходного массива на примере класса "7-42-Н-В-Ми" к законопроекту "О полиции"

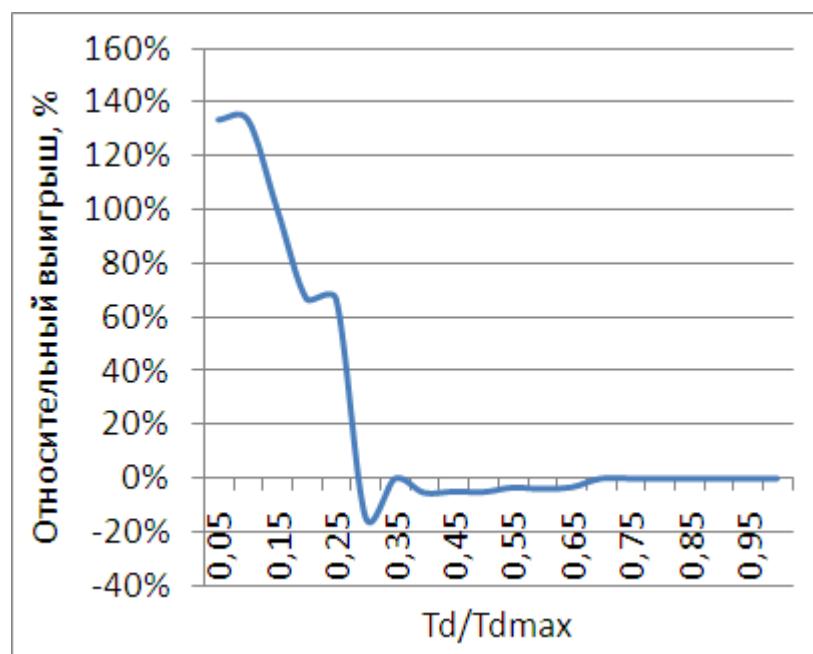


Рисунок 3.16 – Относительный выигрыш в накоплении R1 по сравнению с R1исх на примере класса "7-42-Н-В-Ми" к законопроекту "О полиции"

Оцениваемый алгоритм формирования выборки полезных комментариев показал наилучшие результаты при оценке по среднему приросту полноты вида R1 и R3 при $\Delta t = 0,05 \cdot Td_{\max}$ (рисунки 3.17–3.20).

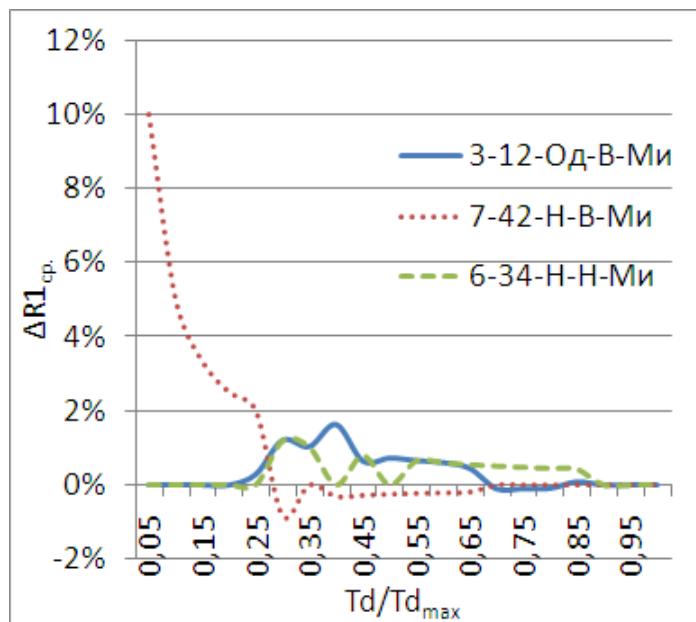


Рисунок 3.17 – Абсолютный выигрыш в среднем приросте полноты вида R1 при $\Delta t = 0,05 \cdot Td_{\max}$ для выборки полезных комментариев к трем классам законопроекта "О полиции" с различным числом объектов комментирования

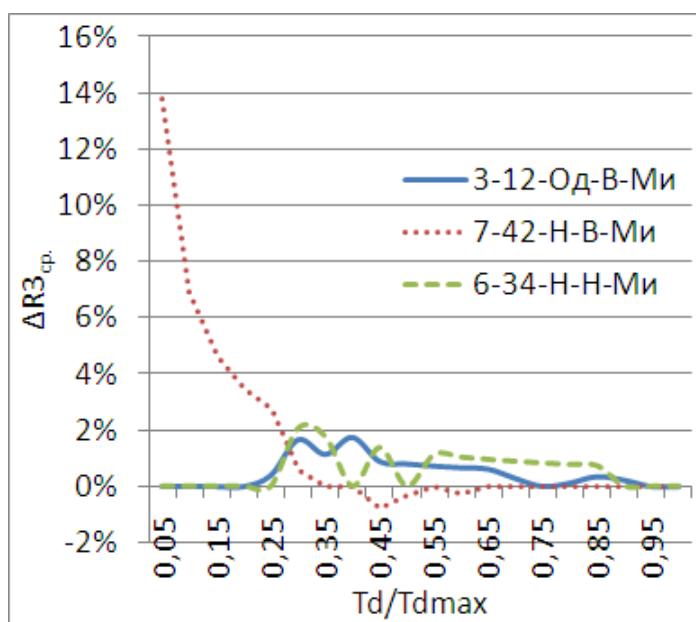


Рисунок 3.18 – Абсолютный выигрыш в среднем приросте полноты вида R3 при $\Delta t = 0,05 \cdot Td_{\max}$ для выборки полезных комментариев к трем классам законопроекта "О полиции" с различным числом объектов комментирования

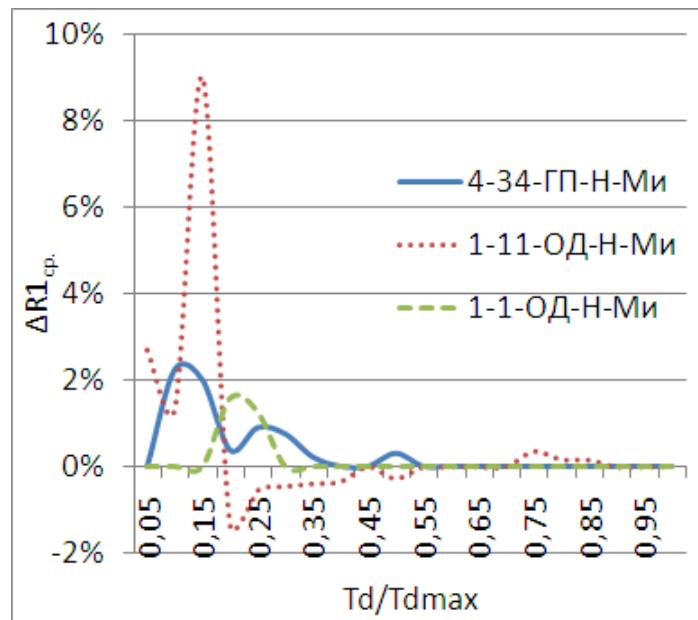


Рисунок 3.19 – Абсолютный выигрыш в среднем приросте полноты вида R1 при $\Delta t = 0,05 \cdot Td_{max}$ для выборки полезных комментариев к трем классам законопроекта "Об образовании" с различным числом объектов
комментирования

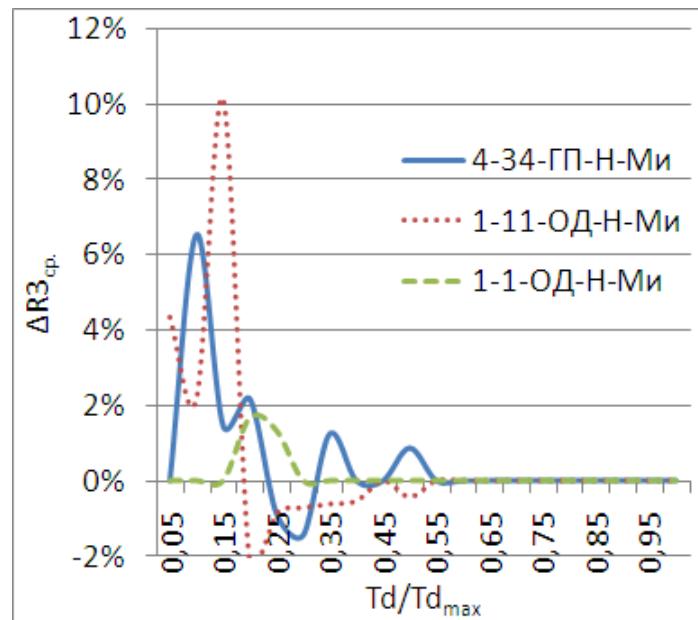


Рисунок 3.20 – Абсолютный выигрыш в среднем приросте полноты вида R3 при $\Delta t = 0,05 \cdot Td_{max}$ для выборки полезных комментариев к трем классам законопроекта "Об образовании" с различным числом объектов
комментирования

Эксперименты показали, что для различных законопроектов и статей с различным количеством объектов комментирования разработанный алгоритм позволяет обеспечить опережающий темп изучения высказываний экспертом, как правило, лишь в первую половину от максимального бюджета времени, что связано с тем, что значительная часть комментариев содержит по одному уникальному высказыванию.

По результатам оценивания эффективности алгоритма формирования выборки полезных комментариев можно сделать следующие выводы:

- разнообразие высказываний в классах комментариев связано с количеством объектов комментирования (пунктов и подпунктов), или мощностью статей законопроекта; распределение статей по количеству объектов комментирования подчинено экспоненциальному закону;
- высказывания распределены по комментариям неравномерно, в проанализированных классах встречаются комментарии с большим числом высказываний, однако доля таких комментариев невелика, значительное число комментариев содержит по одному уникальному высказыванию;
- установлено, что средний прирост полноты зависит от способа ее расчета, в зависимости от учета веса высказываний и идентификации их повторов;
- накопление полноты отбора высказываний во включенных в соответствии с алгоритмом в выборку комментариях в начальную фазу их обработки (до половины бюджета времени, в зависимости от класса) происходит быстрее, при этом на данной фазе полнота отбора высказываний в относительных величинах может значительно (в 2 и более раз) превосходить полноту отбора в неотранжированных по полезности комментариях исходного массива, что объясняется наличием большого числа высказываний в комментариях с высокими рангами;
- направленный отбор комментариев с учетом обобщенного показателя полезности, осуществляемый с использованием разработанного алгоритма, позволяет обеспечить более высокий средний темп прироста полноты отбора высказываний за счет приоритетной обработки комментариев, содержащих большое число высказываний.

3.4 Оценивание свойств алгоритма формирования выборки типичных комментариев

Для алгоритма критическими фрагментами являются блоки 2, 11.

Блок 2. Предусловие выполнения данного блока: $Td_{\max} \neq 0$.

Так как значение Td_{\max} не меняется ни в одном из блоков, предшествующих блоку 2, и по определению не может быть отрицательным числом, то предусловие выполнения алгоритма: $Td_{\max} > 0$.

Блок 11. Так как в блоке 10 на каждом шаге происходит увеличение значения переменной i , то выход из цикла через определенное количество шагов будет всегда осуществлен.

Результат оценивания временной функции сложности алгоритма приведен в таблице 3.5.

Таблица 3.5 – Оценивание временной функции сложности алгоритма формирования выборки типичных комментариев

Номер блока	Коэффициент стоимости	Число раз исполнения для наихудшего случая
1	1	1
2	1	1
3	k^2	1
4	k^2	1
5	1	1
6	k_v^2	Q_v
7	k_v^2	Q_v
8	k_v^2	Q_v
9	1	Q_v
10	1	Q_v
11	1	Q_v
12	1	1

Временная функция сложности алгоритма: $O(3k_v^2Q_v + 2k^2 + 3Q_v + 4)$, где k – число комментариев в классе законопроекта, k_v – число комментариев в кластере v , Q_v – квота для количества типичных комментариев в классе.

Так как временная функция сложности алгоритма представлена полиномом, данный алгоритм относится к классу "легких" алгоритмов.

3.5 Оценивание свойств алгоритма формирования выборки полезных комментариев

Для алгоритма критическими фрагментами являются блоки 2, 4, 10, 12.

Блок 2. Предусловия выполнения данного блока: $p_{\max} \neq 0, l_{\max} \neq 0$.

Так как значения p_{\max} и l_{\max} не меняются ни в одном из блоков, предшествующих блоку 2, и по определению не могут быть отрицательным числом, то предусловия выполнения алгоритма: $p_{\max} > 0, l_{\max} > 0$.

Блок 4. Так как в блоке 3 на каждом шаге происходит увеличение значения переменной j , то выход из цикла через определенное количество шагов будет всегда осуществлен.

Блок 10. Так как в блоке 9 на каждом шаге происходит увеличение значения переменной v , то выход из цикла через определенное количество шагов будет всегда осуществлен.

Блок 12. Предусловия выполнения данного блока: $L_v \neq 0, T1 \neq 0$.

Так как значения L_v и $T1$ не меняются ни в одном из блоков, предшествующих блоку 12, и по определению не могут быть отрицательным числом, то предусловие выполнения алгоритма: $L_v > 0, T1 > 0$.

Вывод: предусловия корректного вычисления алгоритма: $p_{\max} > 0, l_{\max} > 0, L_v > 0, T1 > 0$. Исходя из семантики предметной области, данные условия выполняются всегда, следовательно, алгоритм корректен.

Результат оценивания временной функции сложности алгоритма приведен в таблице 3.6.

Временная функция сложности алгоритма: $O(w^3 + k^2 w + 4w + 3k + s + 4)$, где w – число классов, сформированных комбинацией значимых атрибутов, k – число комментариев к законопроекту, s – число атрибутов комментариев.

Так как временная функция сложности алгоритма представлена полиномом, данный алгоритм относится к классу "легких" алгоритмов.

Таблица 3.6 – Оценивание временной функции сложности алгоритма формирования выборки полезных комментариев

Номер блока	Коэффициент стоимости	Число раз исполнения для наихудшего случая
1	1	1
2	1	k
3	1	k
4	1	k
5	s	1
6	1	1
7	k^2	w
8	1	w
9	1	w
10	1	w
11	w	1
12	w^3	1
13	1	1
14	1	1

Выводы по 3 разделу

1. Разнообразие высказываний в классах комментариев связано с количеством объектов комментирования (пунктов и подпунктов), или мощностью статей законопроекта; распределение статей по количеству объектов комментирования подчинено экспоненциальному закону.
2. Высказывания распределены по комментариям неравномерно, в проанализированных классах встречаются как комментарии с большим числом высказываний, так и высказывания, присутствующие в нескольких комментариях, однако доля таких комментариев и высказываний невелика, значительное число комментариев содержит по одному уникальному высказыванию.
3. Модель обеспечивает выделение высказываний, совпадающих по смыслу с высказываниями, приведенными в обзоре в качестве типичных, выделение типичных высказываний с использованием модели осуществляется в среднем на 20 % быстрее по сравнению с выделением этих же высказываний путем последовательного чтения.
4. Выборка типичных комментариев более эффективна для ознакомления с обобщенными высказываниями, представленными в комментариях

изучаемого класса, а выборка полезных – для ознакомления с разнообразием высказываний в классе комментариев.

5. Средний прирост полноты отбора высказываний зависит от учета при ее расчете веса высказываний и идентификации их повторов.

6. Накопление полноты отбора высказываний во включенных в соответствии с алгоритмами в выборку комментариях в начальную фазу их обработки (до половины бюджета времени, в зависимости от класса) происходит быстрее, при этом на данной фазе полнота отбора высказываний в относительных величинах может значительно (в 1,5 и более раз) превосходить полноту отбора в неотранжированных по типичности либо полезности комментариях исходного массива, что объясняется наличием большого числа высказываний в комментариях с высокими рангами либо типичных высказываний в комментариях с высокими рангами.

7. Использование меры подобия в качестве прототипа семантической метрики и сводного реферата в качестве типичного представителя мнения позволяет на основе оценки подобия группировать комментарии в кластеры, выявлять наличие повторяющихся высказываний и обеспечить путем оценки сходства с типичным представителем мнения в кластере приоритетную обработку комментариев, содержащих часто встречающиеся высказывания.

8. Направленный отбор комментариев с учетом содержательного подобия и сходства с типичным представителем мнения дает возможность обеспечить более высокий средний темп прироста полноты отбора высказываний за счет приоритетной обработки комментариев, содержащих типичные высказывания.

9. Направленный отбор комментариев с учетом обобщенного показателя полезности позволяет обеспечить более высокий средний темп прироста полноты отбора высказываний за счет приоритетной обработки комментариев, содержащих большое число высказываний.

10. Разработанные алгоритмы формирования выборок корректны и относятся к классу алгоритмов с полиномиальной сложностью.

4 НАУЧНО-ТЕХНИЧЕСКИЕ ПРЕДЛОЖЕНИЯ ПО ПОСТРОЕНИЮ И ПРИМЕНЕНИЮ СИСТЕМЫ ОПЕРАТИВНОЙ АНАЛИТИЧЕСКОЙ ОБРАБОТКИ КОММЕНТАРИЕВ

4.1 Структура системы оперативной аналитической обработки комментариев

На практике для реализации разработанной модели и алгоритмов решения задач (1.15) и (1.16) предлагается построить систему оперативной аналитической обработки комментариев, структура которой представлена на рисунке 4.1 [105, 106]. В состав системы входят следующие подсистемы:

- подсистема идентификации комментариев;
- подсистема классификации комментариев;
- блок стандартного OLAP-анализа;
- подсистема формирования выборки типичных комментариев;
- подсистема формирования выборки полезных комментариев.

Рассмотрим назначение, функции блоков и подходы к реализации блоков данных подсистем.

Подсистема идентификации комментариев предназначена для разделения сводного файла с комментариями на отдельные файлы комментариев и выявления содержащихся в сводном файле характеристик комментариев ("служебной" информации).

В блоке выделения комментария методом лексического анализа [107] осуществляется поиск начала и окончания комментария, отделяется "служебная" информация, производится присвоение комментарию идентификатора и сохранение комментария в отдельный файл.

В основе реализации блока лежит применение функций работы со строками: чтения строк из текстового файла, поиска подстроки в строке, записи строк в текстовый файл (рисунок 4.2) [108]. Для функционирования

блока необходимо задать путь к файлу с комментариями, а также выходной каталог для размещения файлов (рисунок 4.3).

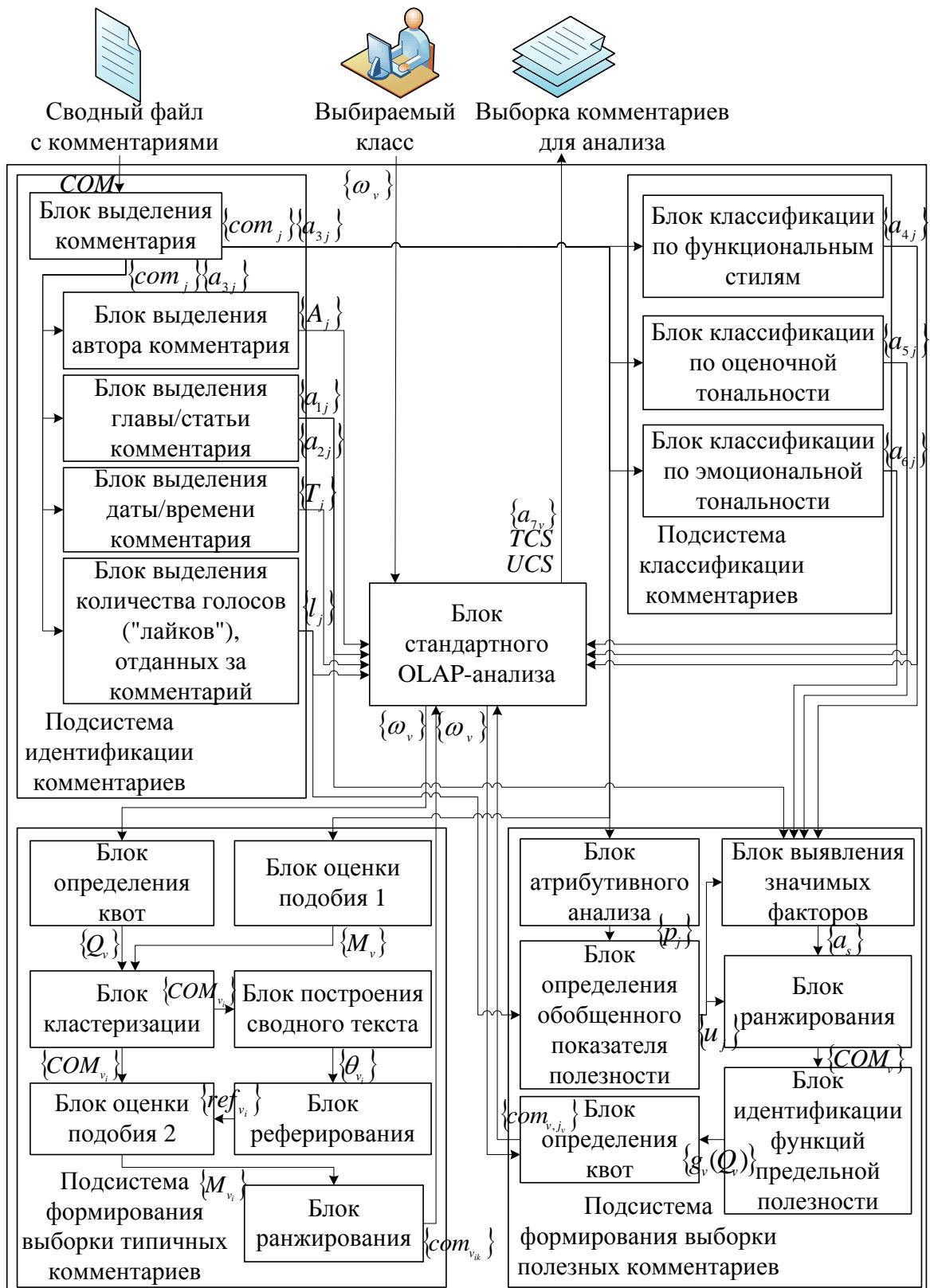


Рисунок 4.1 – Структура системы оперативной аналитической обработки комментариев

```

i:=1;
repeat
  readln(f1,s);
  if pos('К главе',s)>0 then if pos('статье',s)>0 then
begin
  assignfile(f2,inttostr(i)+' '+s+'.txt');
  writeln(filelist,inttostr(i)+' '+s+'.txt');
  rewrite(f2);
  while pos('+',s)<>1 do
begin
  readln(f1,s);
  if pos('+',s)<>1 then writeln(f2,s);
end;
closefile(f2);
label2.caption:=inttostr(i);
progressbar1.position:=i;
inc(i);
end;
until eof(f1);

```

Рисунок 4.2 – Фрагмент программного кода блока выделения
комментария

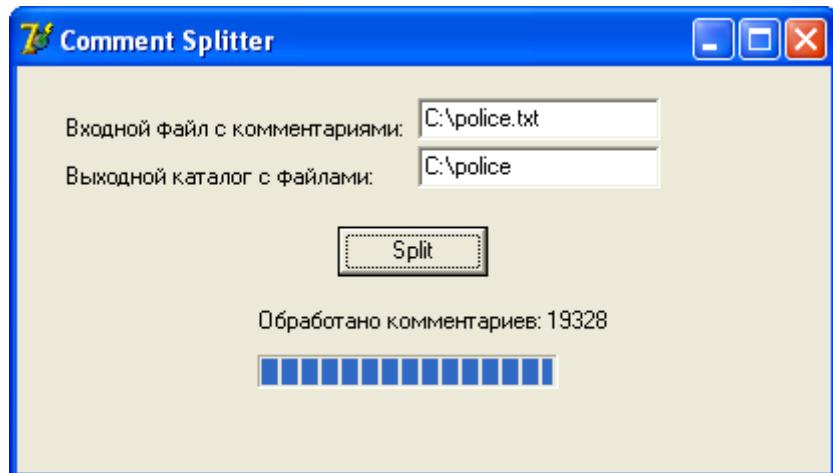


Рисунок 4.3 – Интерфейс блока выделения комментария

В блоке *выделения автора комментария* методом лексического анализа выделяются сведения об авторе (или авторах) комментария и заносятся в таблицу параметров.

В блоке *выделения главы/статьи комментария* методом лексического анализа выделяются сведения о главе и статье комментария и заносятся в таблицу параметров.

В *блоке выделения даты/времени комментария* методом лексического анализа выделяются сведения о дате и времени комментария и заносятся в таблицу параметров.

В *блоке выделения количества голосов, отданных за комментарий*, методом лексического анализа выделяются сведения о количестве голосов в поддержку комментария и заносятся в таблицу параметров.

В основе реализации перечисленных блоков выделения параметров комментариев лежит применение функций работы со строками: чтения строк из текстового файла, возврата подстрок заданного размера в начале и конце строки, определения стартового индекса подстроки в стоке, записи строк в текстовый файл (рисунок 4.4).

Для функционирования блоков необходимо задать путь к файлу с комментариями, а также выходной каталог для размещения файла. Результат функционирования подсистемы идентификации комментариев записывается в текстовый файл и представляет собой набор их параметров (таблица 4.1).

Таблица 4.1 – Фрагмент выходных данных подсистемы идентификации комментариев на примере законопроекта "О полиции"

Идентификатор комментария	Автор	Дата	Время	"Лайков"
...
10350	SWAT	09.08.2010	19:35	1
10351	Роман	09.08.2010	19:35	1
10352	Russel	09.08.2010	19:34	2
10353	NVP	09.08.2010	19:33	4
10354	Филиппов Александр	09.08.2010	19:32	1
10355	SDan	09.08.2010	19:32	0
10356	zolt	09.08.2010	19:32	0
10357	Kroshka	09.08.2010	19:31	0
10358	Крайнов Дмит- рий	09.08.2010	19:31	0
10359	Елена Раду	09.08.2010	19:31	0
...

```

i:=1; //основной цикл по выделению параметров комментариев (автор, дата, время,
      число лайков)
repeat
  readln(f1,s);

  if pos('#',s)>1 then //выкусывание автора, даты и времени из строки с # (>1, а
  не >0, т.к. некоторые используют # для списков)
  begin
    s_adtl:=s;
    au:=leftstr(s_adtl,pos(',',s_adtl)-1);
    s_adtl:=rightstr(s_adtl,length(s_adtl)-(length(au)+2)); // "выкусываем" автора

    repeat
      da:=leftstr(s_adtl,pos(',',s_adtl)-1);
      s_adtl:=rightstr(s_adtl,length(s_adtl)-(length(da)+2)); // "выкусываем" дату
      if (pos('августа',da)>0) then da:=leftstr(da,pos(' ',da)-1) + '.08.2010';
      if (pos('сентября',da)>0) then da:=leftstr(da,pos(' ',da)-1) + '.09.2010';
      if (pos('декабря',da)>0) then da:=leftstr(da,pos(' ',da)-1) + '.12.2010';
      if (pos('января',da)>0) then da:=leftstr(da,pos(' ',da)-1) + '.01.2011';
      if (pos('февраля',da)>0) then da:=leftstr(da,pos(' ',da)-1) + '.02.2011';
      if leftstr(rightstr(da,4),3)<>'201' then au:=au+' '+da; // если после запятой не
      // дата, то дополняем автора
    until leftstr(rightstr(da,4),3)='201';

    ti:=leftstr(s_adtl,pos(' ',s_adtl)-1); // "выкусываем" время
  end;

  if pos('К главе',s)>0 then if pos('статье',s)>0 then
  begin
    assignfile(f2,inttostr(i)+' '+s+'.txt');
    writeln(filelist,inttostr(i)+' '+s+'.txt');
    rewrite(f2);
    while pos('+',s)<>1 do
    begin
      readln(f1,s);
      if pos('+',s)<>1 then writeln(f2,s);
    end;
    li:=rightstr(s,length(s)-pos('+',s)); // "выкусывание" числа лайков
    append(f_adtl);
    writeln(f_adtl,inttostr(i)+';'+au+' '+da+' '+ti+' '+li); //запись параметров
    //адтл= в файл
    closefile(f_adtl);

    closefile(f2);
    label2.caption:=inttostr(i);
    progressbar1.position:=i;
    inc(i);
  end;
  until eof(f1);

```

Рисунок 4.4 – Фрагмент программного кода блоков выделения автора, главы/статьи, даты/времени, количества голосов, отданых за комментарий

Подсистема классификации комментариев предназначена для определения оценочных характеристик комментариев с целью разбиения исходного множества комментариев на классы и последующего анализа высказываний в комментариях с одинаковыми значениями классификационных признаков.

В блоках классификации по функциональным стилям, оценочной и эмоциональной тональностям с использованием словарных и статистических методов обработки текстов производится определение функционального стиля, оценочной и эмоциональной тональностей комментария.

В основе реализации блока классификации по функциональным стилям может быть применен подход, реализованный автором совместно с А. А. Овсянниковым [109] и обеспечивающий высокую точность классификации (более 90 %). Сущность данного подхода заключается в учете систематического отличия лингвостатистических профилей разных функциональных стилей, обусловленного различным использованием языковых средств разных уровней языка в текстах разных стилей (рисунок 4.5). Такой подход к определению функционального стиля текста называют структурно-статистическим. Описанию различных вариантов данного подхода посвящен целый ряд отечественных и зарубежных работ [110–113].

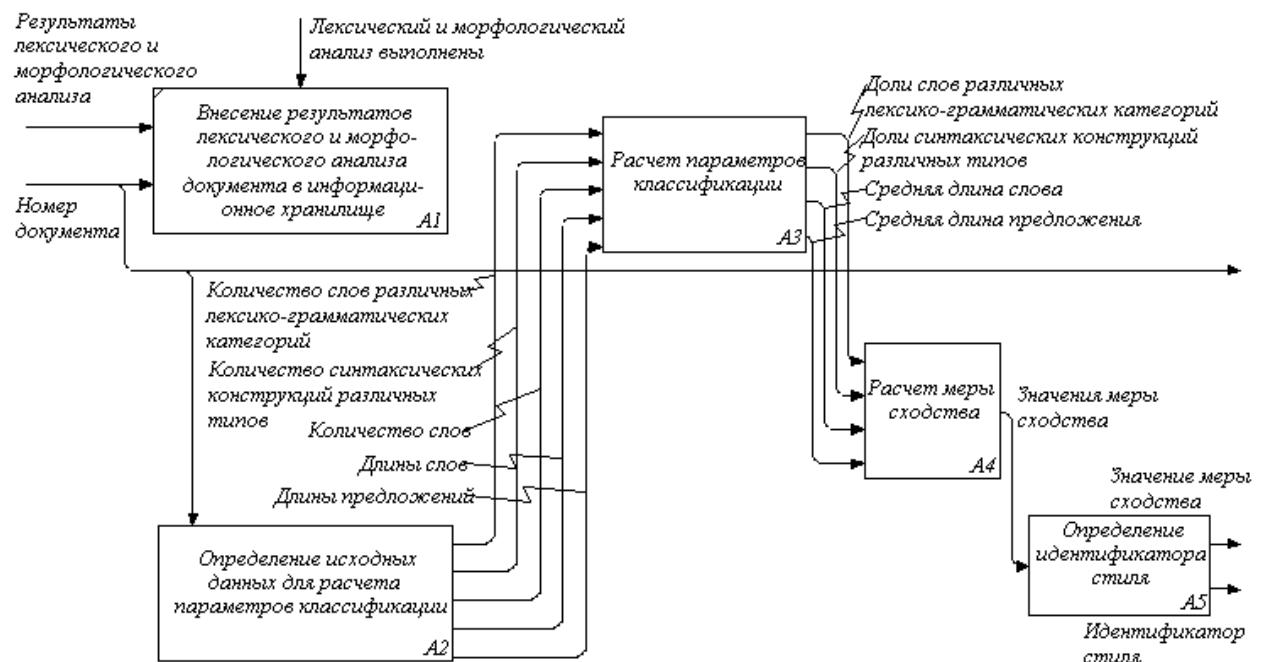


Рисунок 4.5 – Функциональная структура блока классификации по функциональным стилям

Для функционирования блока необходимо:

- выбрать комментарий или указать каталог с комментариями, для которых требуется определить функциональный стиль;
- проводить морфологический анализ комментариев;
- проводить стилистический анализ комментариев.

В результате стилистического анализа в окне результатов расчета признаков отображаются значения признаков, а в окне результата – функциональный стиль комментария (рисунок 4.6). При обработке потока комментариев результат анализа записывается в файл.

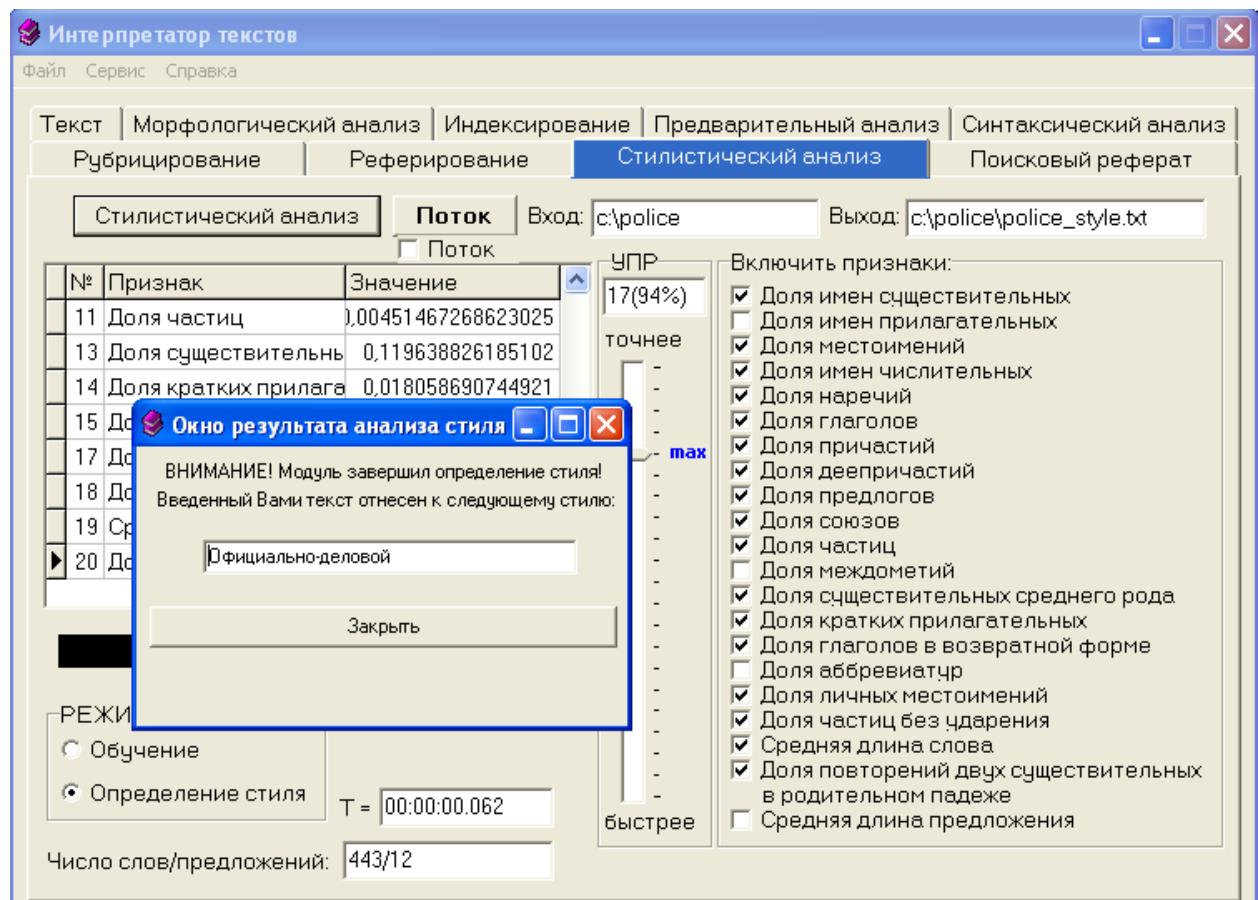


Рисунок 4.6 – Интерфейс блока классификации по функциональным стилям

В основе реализации блока классификации по оценочной тональности может быть применен подход, реализованный А. А. Кукушкиным и И. С. Полянским [114] и обеспечивающий приемлемую точность классификации (более 70 %). Сущность данного подхода заключается в комбинировании оценок

тональности блоков текста более низкого уровня для определения тональности блоков текста более высокого уровня (рисунок 4.7). На нижнем уровне определение тональности существительных, прилагательных и глаголов производится с использованием семантического словаря тональностей, на последующих – с использованием словарей составных индикаторов тональностей и матриц комбинирования оценок.



Рисунок 4.7 – Алгоритм функционирования блока определения оценочной тональности

Для функционирования блока необходимо задать путь к файлам с комментариями, а также выходной каталог. Результат функционирования блока записывается в текстовый файл.

Результат функционирования подсистемы классификации комментариев записывается в текстовый файл и представляет собой набор их параметров (таблица 4.2).

Таблица 4.2 – Фрагмент выходных данных подсистемы классификации комментариев на примере законопроекта "О полиции"

Идентификатор комментария	Функциональный стиль	Эмоциональная тональность	Оценочная тональность
...
10350	научный	высокая	минорная
10351	научный	высокая	минорная
10352	официально-деловой	низкая	минорная
10353	художественно-беллетристический	высокая	минорная
10354	газетно-публицистический	низкая	мажорная
10355	газетно-публицистический	высокая	минорная
10356	научный	высокая	минорная
10357	научный	низкая	минорная
10358	научный	высокая	минорная
10359	газетно-публицистический	высокая	минорная
...

Блок стандартного OLAP-анализа предназначен для оперативной аналитической обработки комментариев с использованием сведений об их разбиении и распределении по классам с целью установления очередности обработки в зависимости от интенсивности мнений и предпочтений аналитика.

В основе реализации блока могут быть применены библиотеки, поддерживающие стандартный набор OLAP-операций: сечение, вращение, погружение, всплытие для количественных данных.

Для функционирования блока необходимо задать пути к таблицам с выходными данными подсистем идентификации и классификации комментариев. Интерфейс блока должен предусматривать возможность осуществления операций визуального анализа данных (рисунок 4.8).

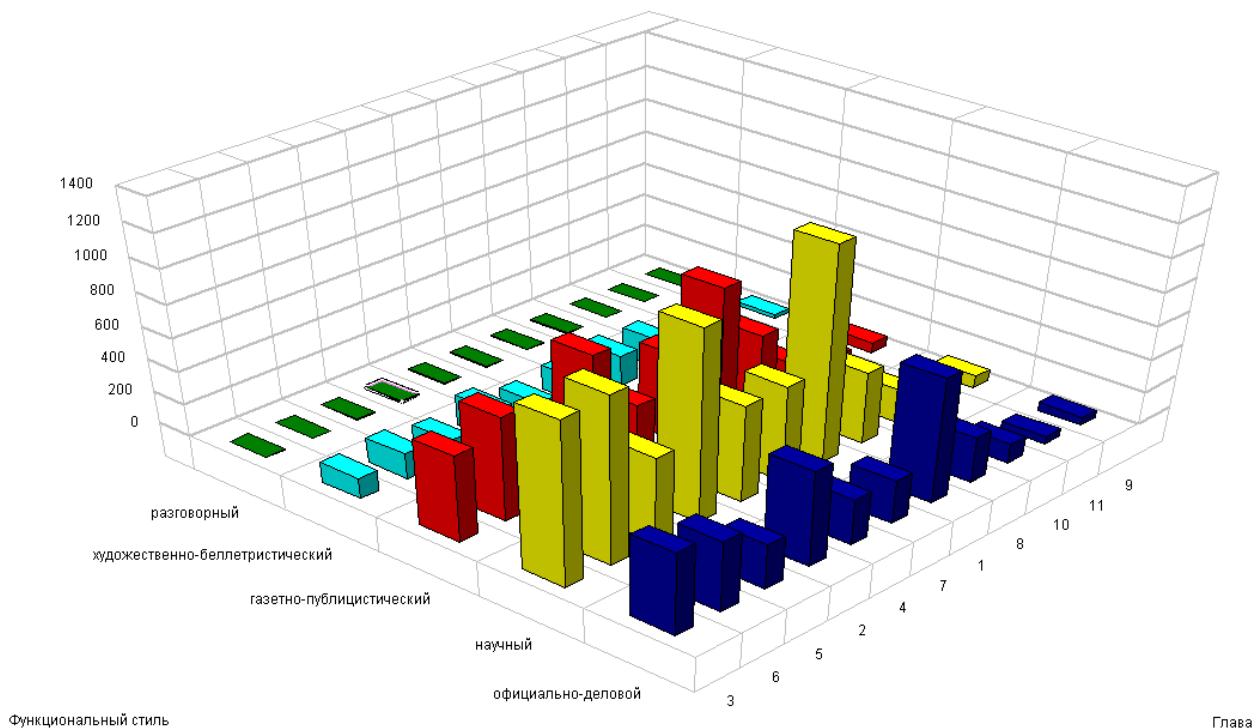


Рисунок 4.8 – Интерфейс блока стандартного OLAP-анализа на примере пакета Cognos PowerPlay [115]

Подсистема формирования выборки типичных комментариев предназначена для детализации выбираемого класса комментариев на основе решения системы (1.15) и учитывает неравномерность распределения высказываний в комментариях посредством определения содержательного подобия комментариев и сходства с типичным представителем мнения.

В блоке *определения квот* с учетом бюджета времени определяется мощность выборки.

В основе реализации блока лежит расчетная процедура. Для функционирования блока необходимо задать директивное время, выделенное для анализа комментариев.

В блоке *оценки подобия 1* с использованием выбранной модели подобия строится матрица подобия комментариев класса, на основании которой формируется матрица различий.

В основе реализации блока может быть применен подход к оценке подобия текстов, предложенный А. Ю. Бородащенко [75, 116]. Сущность данного подхода заключается в многоаспектном ранжировании сравниваемых документов на основе анализа их структурных (выявление связей между ключевыми темами документов), контекстных (сравнение цепочек слов) и вне контекстных (анализ связей между документами) элементов. Основу алгоритмов оценки подобия текстов в соответствии с данным подходом составляют графовая, марковская и гипертекстовая модели "семантической" фильтрации текстов. Функциональная структура блока оценки подобия представлена на рисунке 4.9.

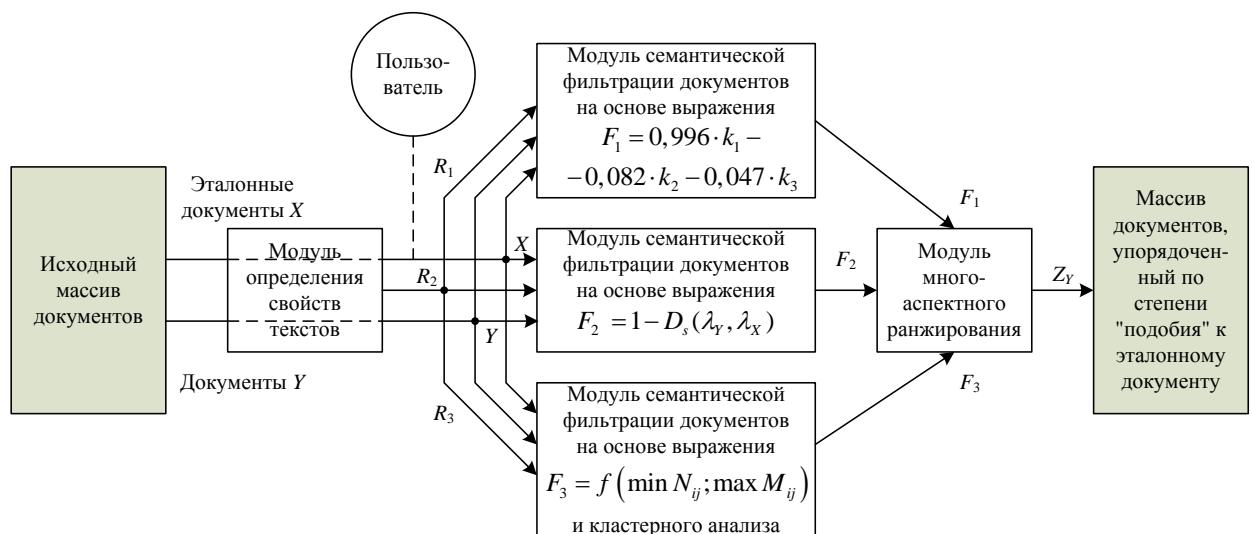


Рисунок 4.9 – Функциональная структура блока оценки подобия

Упрощенный вариант построения блока оценки подобия в случае общей темы и независимых документов может быть реализован с использованием модуля "семантической" фильтрации на основе марковской модели.

Для функционирования блока необходимо задать путь к файлам с комментариями. Результатом работы блока является матрица подобия класса комментариев, сохраняемая в формате ".csv". На основе полученной матрицы подобия формируется матрица различий.

В *блоке кластеризации* полученная матрица различий применяется для кластеризации комментариев на число кластеров, соответствующее мощности выборки.

В основе реализации блока могут быть применены библиотеки, поддерживающие кластеризацию с использованием метода иерархического объединения.

Для функционирования блока необходимо указать число кластеров (квоту для изучаемого класса комментариев) и путь к файлу с матрицей различий.

В *блоке построения сводного текста* производится объединение комментариев кластера в один файл (рисунок 4.10).

```

repeat                                //ОСНОВНОЙ ЦИКЛ
  readln(f_spis, spis);    //считываем очередной идентификатор для объединения
  reset(f_list);           //открываем список файлов комментариев

  repeat //ищем по всему списку файлов комментариев

    readln(f_list, list);           //считываем имя файла комментария

    m:=['0','1','2','3','4','5','6','7','8','9']; // 
    ide:='';                                //
    l:=1; num:=true;                         //
    for k := 1 to length(list) do begin      //
      if list[k] in m then num:=true;        //
      if list[k] in m then                  //извлечение идентификатора
      case l of                         //из имени файла комментария
        1: ide:=ide+list[k];                //
      end;                                //
      if not (list[k] in m) and (num=true) then inc(l); //
      if not (list[k] in m) then num:=false; //
    end;

    if ide=spis then                  //сравнение идентификатора с взятым из списка
    begin                            // ... и если они совпали, то ...
      assignfile(f_comm, list);
      reset(f_comm); append(f_out);
      repeat
        readln(f_comm, comm);    // ... добавляем содержимое комментария
        writeln(f_out, comm);   //      в общий выходной файл!!!
      until eof(f_comm);
      closefile(f_comm); closefile(f_out);
    end;

    until eof(f_list);   //до конца общего списка файлов комментариев

    closefile(f_list);
    progressbar1.Position:=ind; inc(ind);
  until eof(f_spis); //до конца списка комментариев для объединения

```

Рисунок 4.10 – Фрагмент кода блока построения сводного текста

В блоке реферирования с использованием выбранного метода реферирования производится построение сводного реферата. Режим реферирования должен обеспечивать наилучшее среднее подобие полученного сводного реферата на множество текстов, для которых он строится.

В основу реализации блока может быть положен подход на основе методов статистического автореферирования. В настоящее время одним из перспективных и научно обоснованных методов данного класса является метод сканирующих статистик [97]. Его сущность заключается в выявлении на основе анализа отклонений от равномерности в распределении лексических единиц кластеров (сгущений) отдельных типов единиц, которые можно трактовать как сверхфразовые единства, определяющие тематику текста. Функциональная структура блока реферирования на основе метода сканирующих статистик представлена на рисунке 4.11.



Рисунок 4.11 – Функциональная структура блока реферирования

Для функционирования блока указывают путь к сводному тексту кластера, задают режим реферирования и определяют путь для сохранения реферата.

В *блоке оценки подобия 2* с использованием модели подобия, аналогичной модели в *блоке оценки подобия 1*, определяется комментарий кластера, обладающий наилучшим подобием с полученным сводным рефератом кластера (типичный комментарий).

Подходы к реализации и обеспечению функционирования блока аналогичны блоку оценки подобия 1.

В *блоке ранжирования* определяется порядок ознакомления с комментариями выборки методом установления последовательности включения комментариев в выборку с ростом бюджета времени.

В основе реализации блока может быть применена процедура, направленная на обработку результатов кластерного анализа для установления последовательности разбиения исследуемого класса комментариев на кластеры, т. е. процедуры, обратной иерархическому объединению, в ходе которой очередьность вновь образуемых кластеров определяет ранги новых типичных комментариев данных кластеров. При возникновении редкой ситуации появления двух новых типичных комментариев более высокий ранг целесообразно присвоить типичному комментарию кластера, образованного с меньшим расстоянием связывания.

Для функционирования блока необходимо указать порядок связывания комментариев в кластеры в ходе выполнения процедуры иерархического объединения и список типичных комментариев в кластерах. Результатом функционирования блока является ранжированный список идентификаторов типичных комментариев.

Результатом функционирования подсистемы формирования выборки типичных комментариев является упорядоченное множество типичных комментариев определенной мощности.

Подсистема формирования выборки полезных комментариев предназначена для детализации выбиравшегося класса комментариев на основе решения системы (1.16) и учитывает неравномерность распределения высказываний.

ний в комментариях посредством определения атрибутов конструктивных предложений и голосов интернет-пользователей.

В блоке *атрибутивного анализа* методом лексического анализа с применением словарей синонимичных атрибутов трех типов (атрибуты конструктивных предложений по добавлению, изменению, удалению фрагментов статей законопроекта) производится выявление и подсчет содержащихся в комментарии конструктивных предложений.

В основе реализации данного блока может быть применен подход, реализованный автором совместно с А. А. Кукушкиным и Е. А. Шевченко [117]. Сущность данного подхода заключается в индексировании комментариев, содержащихся в текстовом файле, по заданным атрибутам, размещенным в специально созданной базе знаний, основу которой составляют текстовые файлы, содержащие списки синонимичных атрибутов терминов "дополнить", "изменить", "удалить". Файлы имеют следующую структуру:

```

<файл>:= <список атрибутов>;
<список атрибутов>:={<атрибут><служебный символ>};
<атрибут>:=синонимы терминов "дополнить" | "изменить" | "удалить";
<служебный символ>:= запятая.

```

Для найденных с использованием методов лексического анализа в тексте комментариев атрибутов осуществляются:

- цветографическая разметка фрагментов текста комментария в соответствии с заданными атрибутами;
- представление пользователю размеченных комментариев по классам;
- подсчет и представление статистических показателей атрибутивного анализа;
- вывод данных, полученных в результате проведения атрибутивного анализа, в файл формата ".csv".

Для функционирования блока необходимо:

- а) указать каталог с файлом комментариев, для которых требуется провести атрибутивный анализ;

- б) провести разметку комментариев;
- в) провести подсчет статистики атрибутов конструктивных предложений в комментариях.

В результате атрибутивного анализа в окне результатов анализа отображаются размеченные комментарии, а в окне статистики – статистика конструктивных предложений (рисунок 4.12). В результате обработки комментариев результат атрибутивного анализа записывается в файл.

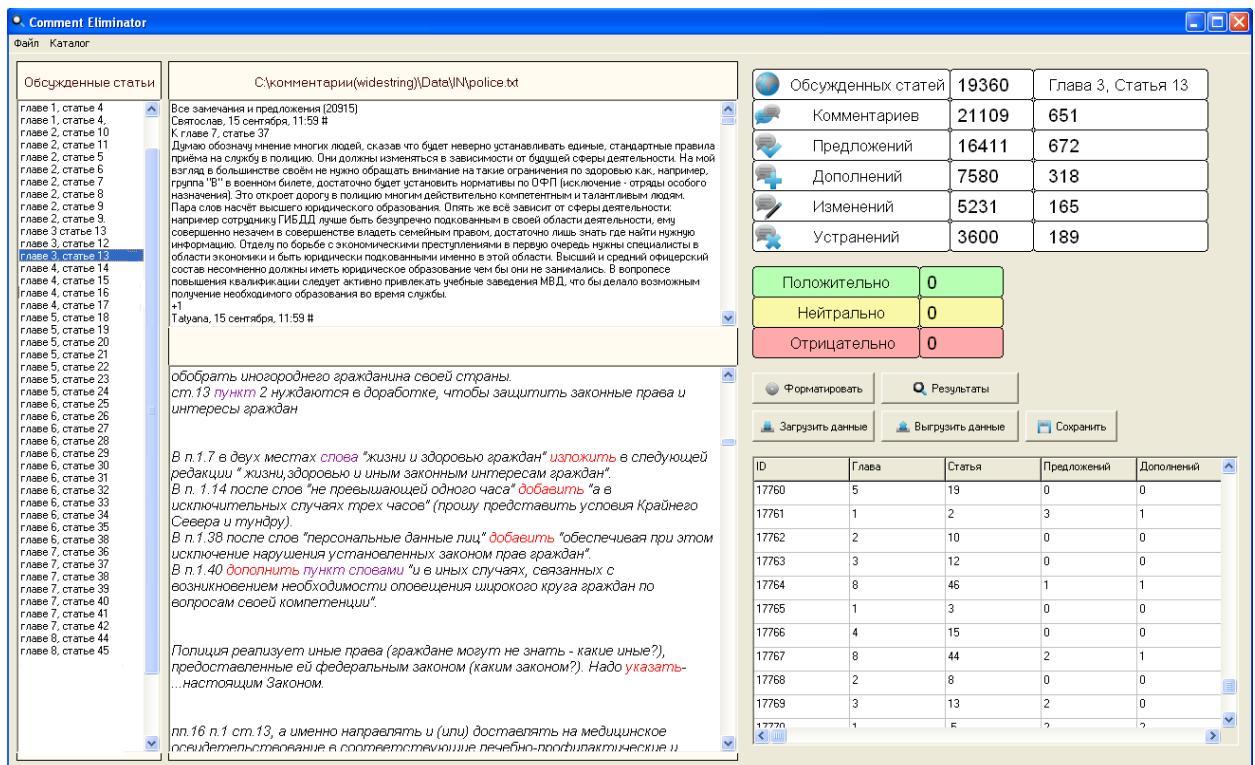


Рисунок 4.12 – Интерфейс блока атрибутивного анализа

В блоке определения обобщенного показателя полезности производится расчет данного показателя посредством свертки нормированных значений числа конструктивных предложений в комментарии и числа голосов, отданых за комментарий, с учетом распределений данных параметров.

В основе реализации блока лежит расчетная процедура. Для функционирования блока требуется указать путь к файлам с данными о количестве голосов, отданых за комментарии, и количестве конструктивных предложений в комментариях, полученных в результате атрибутивного анализа. Ре-

зультатом функционирования блока являются значения обобщенного показателя полезности комментариев.

В блоке *выявления значимых факторов* методом дисперсионного анализа выявляются классификационные признаки, оказывающие значимое влияние на значение обобщенного показателя полезности.

В основу реализации блока могут быть положены библиотеки, обеспечивающие проведение дисперсионного анализа.

Для функционирования блока необходимы выходные данные подсистемы классификации и результаты расчета обобщенного показателя полезности. Результатом функционирования блока является набор атрибутов комментариев, оказывающих значимое влияние на обобщенный показатель полезности комментариев.

В блоке *ранжирования* для каждой комбинации значимых классификационных признаков производится сортировка комментариев по убыванию значения обобщенного показателя полезности.

В основу реализации блока могут быть положены библиотеки, обеспечивающие проведение сортировки данных.

Для функционирования блока необходимы фрагмент выходных данных подсистемы классификации для классов, оказывающих значимое влияние на полезность комментариев, и результаты расчета обобщенного показателя полезности. Результатом функционирования блока является отсортированная по убыванию значения обобщенного показателя полезности последовательность комментариев для каждой комбинации атрибутов, оказывающих значимое влияние на полезность комментариев.

В блоке *идентификации функций предельной полезности* производятся аппроксимация полученных в блоке ранжирования рядов значений аналитическими (логарифмическими) функциями и фиксация параметров данных функций.

В основе реализации блока могут применяться библиотеки, обеспечивающие нахождение параметров логарифмической аппроксимирующей кривой для ряда значений.

Для функционирования блока необходимы ранжированные последовательности значений полезности комментариев для комбинаций значимых классов. Результатом функционирования блока являются параметры аппроксимирующих кривых для всех комбинаций классов, оказывающих значимое влияние на полезность комментариев.

В блоке *определения квот* с учетом общего бюджета времени, темпа чтения, а также функций предельной полезности, средней длины комментария и ограничения на максимальную мощность выборки для каждой комбинации значимых классов, методом множителей Лагранжа определяется мощность выборки для каждой такой комбинации.

В основе реализации блока могут быть применены библиотеки либо инструментальные средства математических пакетов прикладных программ, поддерживающие нелинейную оптимизацию или решение систем нелинейных уравнений.

Для функционирования блока необходимы параметры функций предельной полезности для комбинаций значимых классов, директивное время, выделенное для анализа комментариев, параметр темпа чтения, данные о мощности и средней длине комментариев классов, сформированных комбинацией значимых атрибутов.

Результатом функционирования подсистемы формирования выборки полезных комментариев является упорядоченное множество полезных комментариев определенной мощности.

Ознакомление с комментариями выбранной комбинации классов производится в пределах данной выборки в последовательности, установленной в блоке ранжирования.

4.2 Методика применения системы оперативной аналитической обработки комментариев

Аналитическую обработку комментариев с использованием предложенной системы рекомендуется проводить путем применения следующей методики (рисунок 4.13) [100, 101].

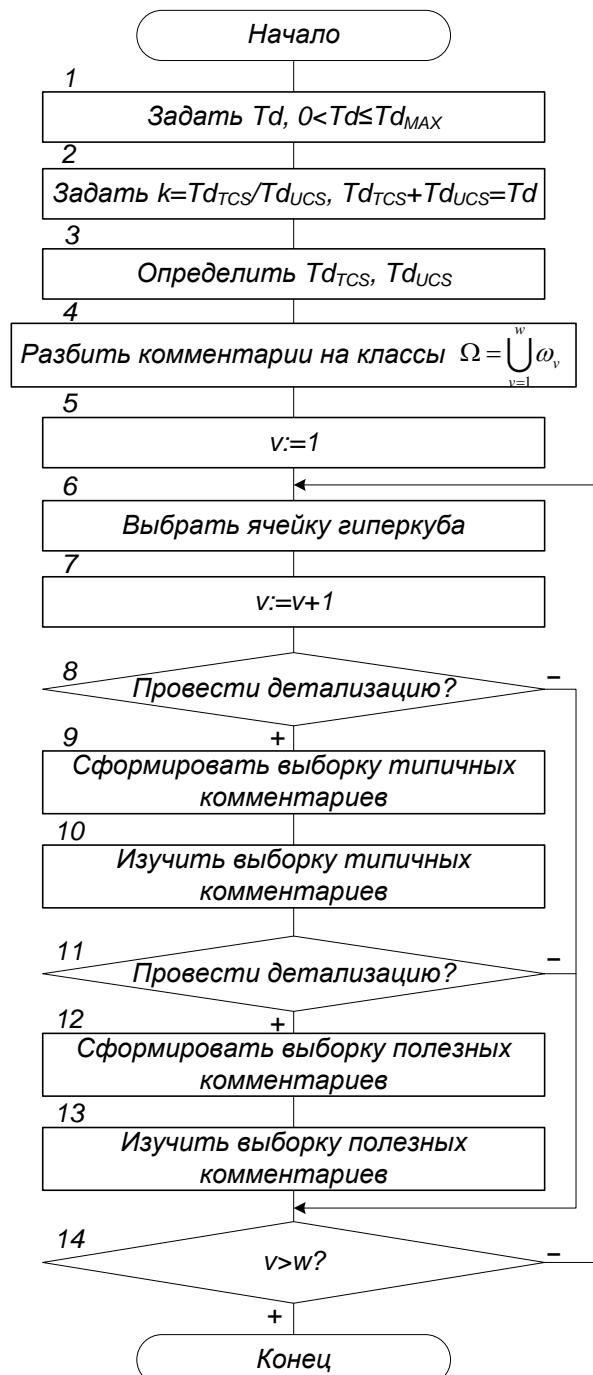


Рисунок 4.13 – Методика применения системы оперативной аналитической обработки комментариев

Методика включает в себя следующие шаги.

На шаге 1 задается директивное время на изучение комментариев. Оно не должно превышать время, требующееся для изучения всех комментариев.

На шаге 2 производится распределение заданного директивного времени между типичными и полезными комментариями путем задания коэффициента k .

На шаге 3 проводится расчет бюджета времени для изучения типичных и полезных комментариев на основе заданных директивного времени и коэффициента k .

На шаге 4 производится классификация комментариев по стилям, эмоциональной и оценочной тональностям, статьям законопроекта.

На шаге 6 выбирается ячейка гиперкуба, т. е. интересующая эксперта комбинация классов.

На шаге 8 принимается решение о проведении детализации с использованием выборки типичных комментариев.

На шаге 9 производится формирование выборки типичных комментариев, алгоритм выполняется каждый раз для выбранной комбинации классов.

На шаге 10 проводится изучение выборки типичных комментариев.

На шаге 11 принимается решение о проведении детализации с использованием выборки полезных комментариев.

На шаге 12 производится формирование выборки полезных комментариев, алгоритм выполняется при первой детализации для всех комбинаций классов, при последующих детализациях производится обращение к результатам работы алгоритма для требуемой комбинации классов.

На шаге 13 проводится изучение выборки полезных комментариев.

Иллюстрация применения методики в ходе аналитической обработки комментариев класса "1-11-Од-Н-Ми" (глава 1, статья 11, стиль – официаль но-деловой, эмоциональная тональность – низкая, оценочная тональность – минорная) к законопроекту "Об образовании" представлена на рисунке 4.14.

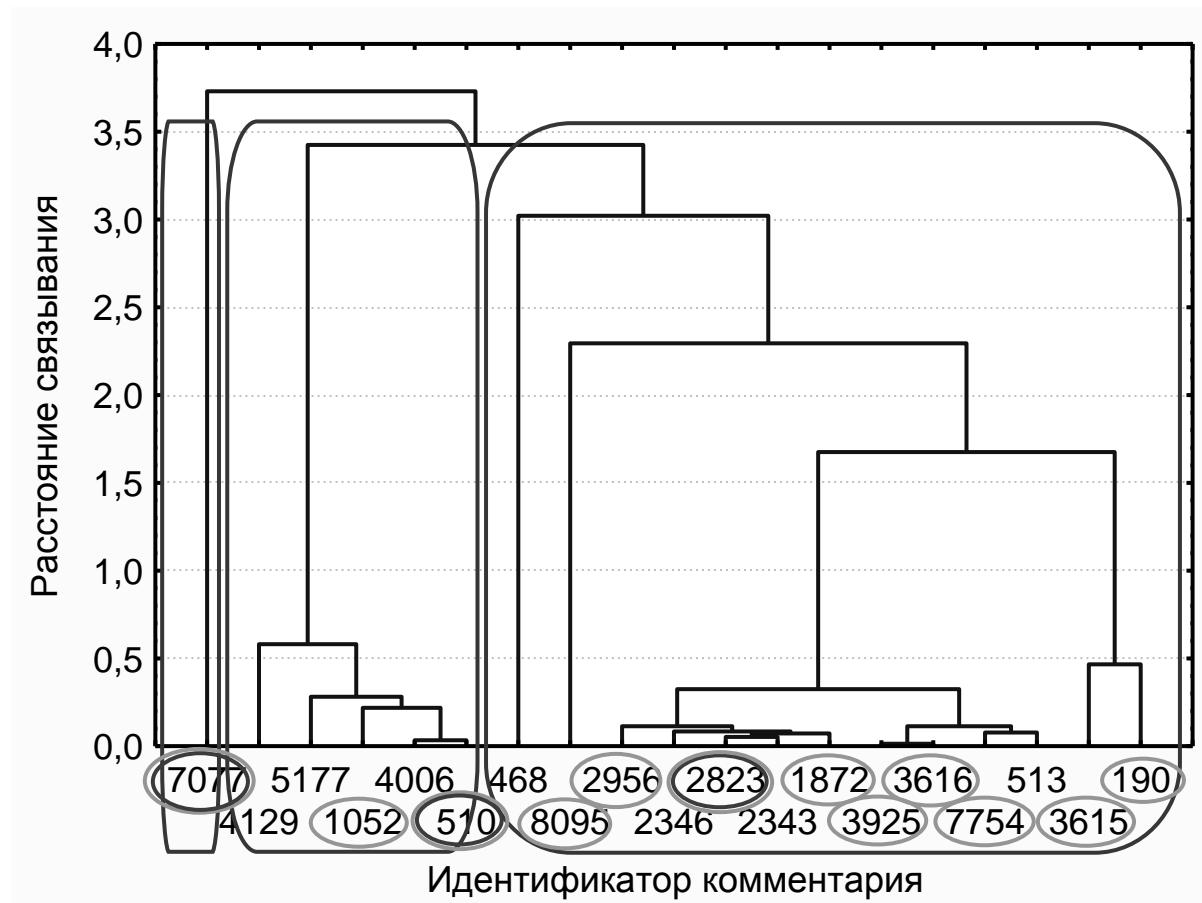


Рисунок 4.14 – Иллюстрация применения методики

В указанном примере детализация проведена при $Td = 0,35 \cdot Td_{\max}$, $k = 0,75$. При данных параметрах эксперту предоставляются для изучения три типичных комментария (идентификаторы обведены темными овалами) и двенадцать полезных комментариев (идентификаторы обведены светлыми овалами). В данном классе комментариев выявлено три обобщенных мнения, представленных типичными комментариями и подкрепленных одним или несколькими полезными комментариями.

Выводы по 4 разделу

1. Реализация системы оперативной аналитической обработки комментариев в соответствии с предложенными подходами позволяет применять на практике разработанные модель и алгоритмы.
2. Основу построения блоков системы составляют методы и алгоритмы, относящиеся к классам Data Mining (англ.: глубинный анализ данных) и Text Mining (англ.: глубинный анализ текстов), что объясняется наличием статистических закономерностей в массиве комментариев и нечисловой природой комментария как отдельно взятого объекта.
3. Функционирование системы осуществляется в автоматизированном режиме: выбор классов комментариев для анализа и изучение полученных выборок комментариев осуществляются экспертом.
4. Блок стандартного OLAP-анализа совместно с подсистемами формирования выборок типичных и полезных комментариев обеспечивают выполнение концепции построения системы обработки комментариев за счет последовательного погружения и ознакомления с сущностью предложений по корректировке законопроекта по принципу от общего представления к детальному.
5. Методика применения системы оперативной аналитической обработки обеспечивает возможность варьирования распределением имеющегося директивного времени между типичными и полезными комментариями и выбора глубины детализации.
6. Разработанные структура и методика применения системы оперативной аналитической обработки комментариев позволяют производить их обработку с учетом неравномерного распределения высказываний и сходства с типичным представителем мнения при заданном бюджете времени.

ЗАКЛЮЧЕНИЕ

В настоящее время в соответствии с Постановлением Правительства РФ от 25.08.2012 № 851 разрабатываемые законопроекты должны проходить процедуру общественного обсуждения на интернет-портале regulation.gov.ru. Результатом такой процедуры является массив комментариев, содержащих предложения интернет-пользователей по корректировке законопроекта.

Одним из подходов к моделированию процедур обработки мнений может быть подход, учитывающий многомерность представления мнения. В работе поставлена и решена научная задача по разработке модели эффективной обработки комментариев с использованием оперативной аналитической обработки текстов, обеспечивающей поддержку рекурсивной процедуры понимания содержания комментариев экспертом на основе дедуктивной схемы.

По итогам решения данной задачи получены следующие результаты:

1. Проведен анализ процесса обработки комментариев, полученных в ходе общественного обсуждения законопроектов. Установлено противоречие между потребностью в эффективных моделях обработки комментариев и отсутствием качественной модели обработки комментариев. Представлен вариант разрешения противоречия.

2. Разработана модель оперативной аналитической обработки комментариев, отличающаяся от известных способом формирования гиперкуба на основе многомерного представления мнения, а также набором операций погружения в текстовые данные, учитывающих неравномерность распределения высказываний в комментариях интернет-пользователей.

3. Синтезирован алгоритм формирования выборки типичных комментариев, учитывающий содержательное подобие комментариев и сходство с типичным представителем мнения для приоритетного отбора комментариев.

4. Синтезирован алгоритм формирования выборки полезных комментариев, учитывающий атрибуты конструктивных предложений и предпочтения интернет-пользователей для приоритетного отбора комментариев.

5. Предложены структура и методика применения системы оперативной аналитической обработки комментариев, позволяющие производить их обработку с учетом неравномерного распределения высказываний и сходства с типичным представителем мнения при заданном бюджете времени.

В результате решения научной задачи обеспечено повышение темпа изучения совокупности высказываний в ходе обработки результатов общественного обсуждения законопроектов:

1. При небольшом бюджете времени, выделенном на обработку (до половины бюджета времени, в зависимости от класса), предложенный подход позволяет в 1,5 и более раз увеличить количество исследованных экспертом высказываний, что обеспечивает эффективное использование выделенного временного ресурса на изучение комментариев.

2. Разработанная модель обеспечивает изучение высказываний, совпадающих по смыслу с высказываниями, приведенными в обзоре в качестве типичных, в среднем на 20 % от максимального директивного времени быстрее по сравнению с изучением этих же высказываний путем последовательного чтения, что ускоряет формирование обобщенного представления о поступивших предложениях.

Направлением дальнейших исследований, по мнению автора, является адаптация разработанных модели и алгоритмов к обработке результатов различных типов обсуждений, в ходе которых мнения выражаются либо могут быть представлены в текстовом виде.

Полученные результаты изложены в 8 печатных работах, из них 3 – в журналах из Перечня ВАК Минобрнауки РФ [80, 94, 105], по материалам исследования получено 2 свидетельства о регистрации программ для ЭВМ в Роспатенте.

СПИСОК СОКРАЩЕНИЙ И УСЛОВНЫХ ОБОЗНАЧЕНИЙ

ЗП	– законопроект
НПА	– нормативный правовой акт
МК	– массив комментариев
ОАОТ	– оперативная аналитическая обработка текстов
СС	– семантическая сеть
ЛЛМ	– логико-лингвистическая модель
ИПТ	– информационно-поисковый тезаурус
ФМ	– фреймовая модель
ОЛ	– онтология
АОТ	– автоматическая обработка текстов
OLAP	– (online analytical processing) оперативная аналитическая обработка данных
TCS	– (typical comments sample) выборка типичных комментариев
UCS	– (utility comments sample) выборка полезных комментариев
ВТК	– выборка типичных комментариев
ВПК	– выборка полезных комментариев

СЛОВАРЬ ТЕРМИНОВ

Обсуждение законопроектов – процедура выражения мнений в отношении содержания статей законопроектов в форме текстовых комментариев, создаваемых с использованием интерфейса сайта общественного обсуждения законопроектов.

Сайт общественного обсуждения законопроектов – интернет-портал или раздел интернет-портала, средства которого обеспечивают процедуры создания, накопления, хранения и обработки комментариев в ходе обсуждения законопроектов.

Респондент – интернет-пользователь, прошедший регистрацию на сайте общественного обсуждения законопроектов и получивший возможность комментировать законопроект и голосовать за понравившиеся комментарии.

Мнение – комплексное отношение к законопроекту, сформировавшееся в сознании респондента по итогам его изучения.

Конструктивное предложение – совокупность идей в сознании респондента по корректировке законопроекта.

Высказывание – совокупность предложений естественного языка, выражающих законченную мысль по корректировке законопроекта.

Комментарий – текст, содержащий совокупность высказываний и соответствующий требованиям документа "Некоторые замечания, рекомендации и правила общественного обсуждения проектов федеральных законов".

Атрибут комментария – смысловой признак комментария, характеризующий его содержание или оформление (тема, тональность, стиль и т. д.).

Массив комментариев – множество комментариев, относящихся к статьям одного законопроекта, упорядоченное в порядке поступления от респондентов.

Эксперт – член комиссии по изучению поступивших комментариев и подготовке аналитического обзора.

Класс комментариев – множество комментариев с одинаковыми значениями атрибутов.

Типичность высказывания – количество комментариев в классе комментариев, содержащих данное высказывание.

Выборка типичных комментариев класса – подмножество комментариев изучаемого экспертом класса с максимальной суммарной типичностью содержащихся в них уникальных высказываний, упорядоченное по убыванию типичности высказываний в комментариях.

Выборка типичных комментариев – совокупность выборок типичных комментариев классов.

Полезность комментария – обобщенный показатель, определяемый с использованием свертки числа атрибутов конструктивных предложений в комментарии и числа голосов, отданных за комментарий ("лайков"), характеризующий конструктивность идей по корректировке законопроекта, высказанных в комментарии.

Выборка полезных комментариев класса – подмножество комментариев изучаемого экспертом класса с максимальной суммарной полезностью комментариев, упорядоченное по убыванию их полезности.

Выборка полезных комментариев – совокупность выборок полезных комментариев классов.

Типичное высказывание – высказывание, выделенное экспертами по итогам изучения массива комментариев и обобщения предложений по корректировке законопроекта и включенное в аналитический обзор.

Обобщающая способность выборки – зависимость количества типичных высказываний, вошедших в выборку, от ее мощности.

Накапливающая способность выборки – зависимость количества высказываний, вошедших в выборку, от ее мощности.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. **О порядке раскрытия федеральными органами исполнительной власти информации о подготовке проектов нормативных правовых актов и результатах их общественного обсуждения** : Постановление Правительства РФ от 25 августа 2012 г. № 851 // Российская газета № 5873 – 2012. – 31 августа.
2. **Единый** портал раскрытия информации о подготовке федеральными органами исполнительной власти проектов нормативных правовых актов и результатах их общественного обсуждения [Электронный ресурс] : сайт. – М., 2013. – Режим доступа: <http://www.regulation.gov.ru/>. – Дата обращения: 15.04.2013.
3. **Pang, B.** Opinion Mining and Sentiment Analysis / B. Pang, L. Lee // Foundations and Trends in Information Retrieval, v.2 n.1–2. – 2008. – Pp. 1–135.
4. **Wiebe, J.** Recognizing and organizing opinions expressed in the world press / J. Wiebe, E. Breck, C. Buckley et al. // Proceedings of the AAAI Spring Symposium on New Directions in Question Answering. – 2003.
5. **Cardie, C.** Combining low-level and summary representations of opinions for multi-perspective question answering / C. Cardie, J. Wiebe, T. Wilson et al. // Proceedings of the AAAI Spring Symposium on New Directions in Question Answering. – 2003. – Pp. 20–27.
6. **Seki, Y.** Analysis of multi-document viewpoint summarization using multi-dimensional genres / Y. Seki, K. Eguchi, N. Kando // Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications. – 2004. – Pp. 142–145.
7. **Carenini, G.** Multi-document summarization of evaluative text / G. Carenini, R. Ng, A. Pauls // Proceedings of the European Chapter of the Association for Computational Linguistics (EACL). – 2006. – Pp. 305–312.
8. **Beineke, P.** Exploring sentiment summarization / P. Beineke, T. Hastie, C. Manning et al. // Proceedings of the AAAI Spring Symposium on Exploring Attitude and Affect in Text, AAAI technical report SS-04-07. – 2004.

9. **Kawai, Y.** Fair news reader: Recommending news articles with different sentiments based on user preference / Y. Kawai, T. Kumamoto, and K. Tanaka // Proceedings of Knowledge-Based Intelligent Information and Engineering Systems (KES), number 4692 in Lecture Notes in Computer Science. – 2007. – Pp. 612–622.
10. **Liu, B.** Web data mining; Exploring hyperlinks, contents, and usage data / B. Liu // Opinion Mining. Springer. – 2006.
11. **Gamon, M.** Pulse: Mining customer opinions from free text / M. Gamon, A. Aue, S. Corston-Oliver et al. // Proceedings of the International Symposium on Intelligent Data Analysis (IDA), number 3646 in Lecture Notes in Computer Science. – 2005. – Pp. 121–132.
12. **Liu, B.** Opinion observer: Analyzing and comparing opinions on the web / B. Liu, M. Hu, J. Cheng // Proceedings of WWW. – 2005.
13. **Gregory, M. L.** User-directed sentiment analysis: Visualizing the affective content of documents / M. L. Gregory, N. Chinchor, P. Whitney et al. // Proceedings of the Workshop on Sentiment and Subjectivity in Text. – Sydney, Australia. – 2006. – Pp. 23–30.
14. **Morinaga, S.** Mining product reputations on the Web / S. Morinaga, K. Yamanishi, K. Tateishi et al. // Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). – 2002. – Pp. 341–349.
15. **Pinch, T. J.** Six degrees of reputation: The use and abuse of online review and recommendation systems / S. David, T. J. Pinch // First Monday. – 2006.
16. **Hu, M.** Mining and summarizing customer reviews / M. Hu, B. Liu // Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD). – 2004. – Pp. 168–177.
17. **Dewally, M.** Reputation, certification, warranties, and information as remedies for seller-buyer information asymmetries: Lessons from the online comic book market / M. Dewally, L. Ederington // Journal of Business. – 2006. – Vol. 79. – Pp. 693–730.

18. **Resnick, P.** Reputation systems / P. Resnick, K. Kuwabara, R. Zeckhauser et al. // Communications of the Association for Computing Machinery (CACM). – 2000. – Vol. 43. – Pp. 45–48.

19. **Mullen, T.** Sentiment analysis using support vector machines with diverse information sources / T. Mullen, N. Collier // Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP). – 2004. – Pp. 412–418.

20. **Codd, E. F.** Providing OLAP to User-Analysts: An IT Mandate [Электронный ресурс] / E. F. Codd, S. B. Codd, C. T. Salley. – Режим доступа: www.minet.uni-jena.de/dbis/lehre/ss2005/sem_dwh/lit/Cod93.pdf. – Дата обращения: 30.03.2011.

21. **Text Cube: Computing IR Measures for Multidimensional Text Database Analysis** [Электронный ресурс] / Cindy Xide Lin, Bolin Ding, Jiawei Han [и др.]. – Режим доступа: www.cs.uiuc.edu/~hanj/pdf/icdm08_xlin.pdf. – Дата обращения: 22.12.2010.

22. **Topic Modeling for OLAP on Multidimensional Text Databases: Topic Cube and its Applications** [Электронный ресурс] / Duo Zhang, ChengXiang Zhai, Jiawei Han [и др.]. – Режим доступа: ti.arc.nasa.gov/m/profile/oza/files/zhzh09.pdf, 2009. – Дата обращения: 22.12.2010.

23. **iNextCube: Information Network-Enhanced Text Cube** [Электронный ресурс] / Yintao Yu, Cindy X. Lin, Yizhou Sun [и др.]. – Режим доступа: www.cs.uiuc.edu/~hanj/pdf/vldb09_yyu.pdf. – Дата обращения: 22.12.2010.

24. **DocCube: multi-dimensional visualisation and exploration of large document sets** [Электронный ресурс] / Josiane Mothe, Claude Chrisment, Bernard Dousset, Joel Alaux. – Режим доступа: citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.101.9344.pdf 2003. – Дата обращения: 15.02.2011.

25. **Keith, S.** Analyzing Large Collections of Electronic Text Using OLAP [Электронный ресурс] / Steven Keith, Owen Kaser, Daniel Lemire. – www.daniel-lemire.com/fr/documents/publications/apics05-web.pdf. – Дата обращения: 22.12.2010.

26. **Lauw, H. W.** TUBE (Text-cUBE) for Discovering Documentary Evidence of Associations among Entities [Электронный ресурс] / Hady W. Lauw, Ee-Peng Lim, HweeHwa Pang. – Режим доступа: opim.wharton.upenn.edu/~sok/papers/l/p824-lauw.pdf. – Дата обращения: 22.12.2010.
27. **Janet, B.** Cube Index: A Text Index Model for Retrieval and Mining [Электронный ресурс] / B. Janet, A. V. Reddy. – Режим доступа: www.ijcaonline.org/journal/number9/pxc387330.pdf. – Дата обращения: 22.12.2010.
28. **Method** for Performing Effective Drill-down Operations in Text Corpus Visualization and Exploration Using Language Model Approaches for Key Phrase Weighting : заявка 2008/0243482 А1 США : МПК G06F 17/27 / Michal Skubacz, Cai-Nicolas Ziegler. – № 11/797632 ; заявл. 04.05.2007 ; опубл. 02.10.2008. – 13 с.
29. **Зыкин, С. В.** Формирование гиперкубового представления данных со списочными компонентами / С. В. Зыкин // Информационные технологии и вычислительные системы. – 2010. – № 4. – С. 38–46.
30. **Ноженкова, Л. Ф.** OLAP-технологии оперативной информационно-аналитической поддержки организационного управления / Л. Ф. Ноженкова, В. В. Шайдуров // Информационные технологии и вычислительные системы. – 2010. – № 2. – С. 15–27.
31. **Паклин, Н. Б.** Бизнес-аналитика: от данных к знаниям (+CD) / Н. Б. Паклин, В. И. Орешков. – СПб. : Питер, 2009. – 624 с.
32. **Об общественном обсуждении** проектов федеральных конституционных законов и федеральных законов : Указ Президента РФ от 9 февраля 2011 г. № 167 // Российская газета № 5405 – 2011. – 11 февраля.
33. **Общественное обсуждение** законопроектов [Электронный ресурс] : сайт. – М., 2012. – Режим доступа: <http://www.zakonoproekt2012.ru/>. – Дата обращения: 12.01.2012.

34. Об утверждении Правил проведения общественного обсуждения проектов федеральных конституционных законов и федеральных законов : Постановление Правительства РФ от 22 февраля 2012 г. № 159 // Российская газета № 5719. – 2012. – 2 марта.

35. Об утверждении правил подготовки нормативных правовых актов федеральных органов исполнительной власти и их государственной регистрации : Постановление Правительства РФ от 13 августа 1997 г. № 1009 (в ред. Постановления Правительства РФ от 18 декабря 2012 № 1334) // Собрание законодательства РФ № 33. – 1997. – 18 августа. – Ст. 3895; Собрание законодательства РФ № 52. – 2012. – 24 декабря. – Ст. 7507.

36. О мерах по совершенствованию подготовки нормативных правовых актов федеральных органов исполнительной власти, устанавливающих не относящиеся к сфере технического регулирования обязательные требования : Постановление Правительства РФ от 2 мая 2012 г. № 421 // Собрание законодательства РФ № 20. – 2012. – 14 мая. – Ст. 2530.

37. Об антикоррупционной экспертизе нормативных правовых актов и проектов нормативных правовых актов : Постановление Правительства РФ от 26 февраля 2010 г. № 96 // Российская газета № 5125. – 2010. – 5 марта.

38. Об экспертизе нормативных правовых актов федеральных органов исполнительной власти в целях выявления в них положений, необоснованно затрудняющих ведение предпринимательской и инвестиционной деятельности : Постановление Правительства РФ от 29 июля 2011 г. № 633 // Российская газета № 5547. – 2011. – 5 августа.

39. О разработке и утверждении административных регламентов исполнения государственных функций и административных регламентов предоставления государственных услуг : Постановление Правительства РФ от 16 мая 2011 г. № 373 // Собрание законодательства РФ № 22. – 2011. – 30 мая. – Ст. 3169.

40. **О порядке проведения федеральными органами исполнительной власти оценки регулирующего воздействия проектов нормативных правовых актов, проектов поправок к проектам федеральных законов и проектов решений Совета Евразийской экономической комиссии, а также о внесении изменений в некоторые акты Правительства Российской Федерации : Постановление Правительства РФ от 17 декабря 2012 г. № 1318 // Собрание законодательства РФ № 52. – 2012. – 24 декабря. – Ст. 7491.**

41. **Об организации предоставления государственных и муниципальных услуг : Федеральный закон от 27 июля 2010 г. № 210-ФЗ // Российская газета № 5247. – 2010. – 30 июля.**

42. **Об утверждении формы сводного отчета о проведении оценки регулирующего воздействия, формы заключения об оценке регулирующего воздействия, методики оценки регулирующего воздействия : приказ Минэкономразвития России от 27.05.2013 г. № 290 // Российская газета № 6160. – 2013. – 21 августа.**

43. Гегель, Г. Ф. Философия права / Г. Ф. Гегель. – М. : Мысль, 1990. – 524 с.

44. Ожегов, С. И. Толковый словарь русского языка: 80 000 слов и фразеологических выражений / С. И. Ожегов, Н. Ю. Шведова. – Российская академия наук. Институт русского языка им. В. В. Виноградова. – 4-е изд., дополненное. – М. : Азбуковник, 1999. – 944 с.

45. Ефремова, Т. Ф. Современный толковый словарь русского языка / Т. Ф. Ефремова. – М. : АСТ : Астрель, 2006.

46. Толковый словарь русского языка : в 4 т. / под ред. Д. Н. Ушакова. – Репринтное издание. – М., 2000.

47. Википедия – свободная энциклопедия [Электронный ресурс] : сайт. – 2013. – Режим доступа: <http://ru.wikipedia.org/>. – Дата обращения: 08.02.2013.

48. Современный толковый словарь. – М. : Издательство "Большая Советская Энциклопедия", 1997.

49. **Агеев, В. Н.** Семиотика / В. Н. Агеев. – М. : Издательство "Весь Мир", 2002. – 256 с.

50. **Большаков, И. А.** Компьютерная лингвистика: модели, ресурсы, приложения [Электронный ресурс] / И. А. Большаков, А. Гельбух. – Мехико, 2004. – 193 с. – Режим доступа: www.gelbukh.com/clbook. – Дата обращения: 15.02.2011.

51. **Большакова, Е. И.** Формализация лексико-синтаксической информации для распознавания регулярных конструкций естественного языка / Е. И. Большакова, Н. Э. Васильева // Программные продукты и системы. – 2008. – № 4 (84). – С. 103–106.

52. **Волкова, И. А.** Введение в компьютерную лингвистику. Практические аспекты создания лингвистических процессоров : учебное пособие для студентов факультета ВМиК МГУ / И. А. Волкова. – М. : Издательский отдел факультета вычислительной математики и кибернетики МГУ им. М. В. Ломоносова, 2006. – 43 с.

53. **Глушков, И. Н.** Методы и модели ситуационного анализа при управлении развитием сложными техническими системами : монография (научное издание) / И. Н. Глушков, Н. В. Латышев, В. Н. Минаев, С. Н. Остапенко. – Тверь : Онли-пресс, 2012. – 624 с.

54. **Кукушкин, А. А.** Системы искусственного интеллекта. В 2 ч. Ч. 2. Системы поддержки принятия решений : учебное пособие / А. А. Кукушкин. – Орёл : Академия ФСО России, 2010. – 228 с.

55. **Шевченко, Н. В.** Основы лингвистики текста : учебное пособие / Н. В. Шевченко. – М. : "Приор-издат", 2003. – 160 с.

56. **Шемакин, Ю. И.** Компьютерная семантика / Ю. И. Шемакин, А. А. Романов. – М. : НОЦ "Школа Китайгородской", 1995. – 344 с.

57. **Sowa, J. F.** Semantic Networks / J. F. Sowa. – In: Encyclopedia of Cognitive Science. – John Wiley & Sons. – 2006.

58. **Russell, S. J.** Artificial intelligence : a modern approach / S. J. Russell, P. Norvig. – Upper Saddle River, N.J. : Prentice Hall. – 2010.

59. **Моделирование языковой деятельности в интеллектуальных системах** : под ред. А. Е. Кибрика и А. С. Нариньяни. – М. : Наука, 1987.
60. **Захаров, В. Н.** Автоматическое формирование визуального представления смыслового содержания документа / В. Н. Захаров, А. А. Хорошилов // Системы и средства информатики. – 2013. – Т. 23. – № 1. – С. 143–158.
61. **Поспелов, Д. А.** Ситуационное управление: теория и практика / Д. А. Поспелов. – М. : Наука. – Гл. ред. физ.-мат. лит., 1986. – 288 с.
62. **Тарасов, В. Б.** Логико-лингвистические модели в искусственном интеллекте: прошлое, настоящее, будущее [Электронный ресурс] / В. Б. Тарасов. – Режим доступа: <http://posp.raai.org/data/posp2005/Tarasov/tarasov.html>. – Дата обращения: 20.09.2011.
63. **Караулов, Ю. Н.** Лингвистическое конструирование и тезаурус литературного языка / Ю. Н. Караулов. – М. : Издательство "Наука", 1981. – 368 с.
64. **Леонтьева, Н. Н.** Автоматическое понимание текстов: системы, модели, ресурсы : учебное пособие для студентов лингвистических факультетов вузов / Н. Н. Леонтьева. – М. : Издательский центр "Академия", 2006. – 304 с.
65. **Минский, М.** Фреймы для представления знаний / М. Минский – М. : Мир, 1979.
66. **Осипов, Г. С.** Лекции по искусственному интеллекту / Г. С. Осипов. – М. : КРАСАНД, 2009. – 272 с.
67. **Осипов, Г. С.** Системы искусственного интеллекта и нейронные сети [Электронный ресурс] / Г. С. Осипов. – Режим доступа: <http://urist.fatal.ru/Book/Glava13/Glava13.htm>. – Дата обращения: 20.09.2011.
68. **Онтологии и тезаурусы: модели, инструменты, приложения** : учебное пособие / Б. В. Добров, В. В. Иванов, Н. В. Лукашевич, В. Д. Соловьев. – М. : Интернет-Университет Информационных Технологий ; БИНОМ. Лаборатория знаний, 2011. – 173 с. : ил. – (Основы информационных технологий).

69. **Ломов, П. А.** Интеграция онтологий с использованием тезауруса для осуществления семантического поиска / П. А. Ломов, М. Г. Шишаев // Информационные технологии и вычислительные системы. – 2009. – № 3. – С. 49–59.
70. **Наместников, А. М.** Перспективы применения технологии семантического web в интеллектуальных хранилищах данных / А. М. Наместников // Программные продукты и системы. – 2008. – № 4 (84). – С. 56–59.
71. **Чохонелидзе, А. Н.** Подход к реализации автоматизированного поиска онтологической информации в источнике / А. Н. Чохонелидзе, Д. В. Богданов // Программные продукты и системы. – 2011. – № 2 (94). – С. 98–100.
72. **Шильников, П. С.** Компьютерная поддержка построения онтологий / П. С. Шильников // Программные продукты и системы. – 2006. – № 2 (74). – С. 50–52.
73. **Маккинни Дж. Ч.** Методология, процедуры и техника социологии // Беккер Г., Босков А. (ред.) Современная социологическая теория в ее преемственности и изменении. – М. : Изд-во иностранной литературы, 1961. – Гл. 7.
74. **Рамеев, О. А.** Методы анализа многомерных данных : учебно-методические материалы / О. А. Рамеев, А. П. Коваленко. – М., 1988.
75. **Бородащенко, А. Ю.** Модели и алгоритмы семантической фильтрации текстовой информации в информационно-аналитических системах предприятия : диссертация на соискание ученой степени кандидата технических наук : 05.13.01 / Бородащенко Антон Юрьевич. – Орел, 2010. – 190 с.
76. **Ефимова, М. Р.** Общая теория статистики : учебник / М. Р. Ефимова, В. Н. Румянцев, Е. В. Петрова. – М. : ИНФРА-М, 2007. – 416 с.
77. **Остапенко, Р. И.** Математические основы психологии / Р. И. Остапенко – Воронеж : ВГПУ, 2010. – 76 с.
78. **Ядов, В. А.** Социологическое исследование: методология, программы, методы / В. А. Ядов. – М. : Наука, 1972. – 266 с.

79. **Дмитриев, А. П.** Организация статистического исследования : учебно-методическое пособие / А. П. Дмитриев. – Пенза, 2005.
80. **Толкунов, А. А.** Модель оперативной аналитической обработки комментариев с операциями детализации на основе выявления типичных и полезных текстов / А. А. Толкунов // Научное обозрение. – 2013. – № 9. – С. 368–371.
81. **Толкунов, А. А.** Оперативная аналитическая обработка мнений в ходе общественного обсуждения законопроектов / А. А. Толкунов // Актуальные проблемы развития технологических систем государственной охраны, специальной связи и специального информационного обеспечения : VIII Всероссийская межведомственная научная конференция : материалы и доклады (Орёл, 13–14 февраля 2013 г.). В 10 ч. Ч. 8. – Орёл : Академия ФСО России, 2013. – 139 с. – С. 86-89.
82. **Толкунов, А. А.** Модель гиперкуба для представления текста в системе понимания и анализа текстовой информации / А. А. Кукушкин, А. А. Толкунов // Проблемы развития технологических систем государственной охраны, специальной связи и специального информационного обеспечения : Седьмая научно-практическая конференция : сборник материалов (Орёл, 3–4 марта 2011 г.). Ч. 4. – Орёл : Академия ФСО России, 2011. – 186 с. – С. 92–95.
83. **Толкунов, А. А.** Применение гиперкуба как модели представления текста при решении информационно-аналитических задач / А. А. Кукушкин, А. А. Толкунов // Проблемы развития технологических систем государственной охраны, специальной связи и специального информационного обеспечения : Седьмая научно-практическая конференция : сборник материалов (Орёл, 3–4 марта 2011 г.). Ч. 4. – Орёл : Академия ФСО России, 2011. – 186 с. – С. 89–92.
84. **Мандель, И. Д.** Кластерный анализ / И. Д. Мандель. – М. : Финансы и статистика, 1988. – 177 с.
85. **Гухман, А. А.** Введение в теорию подобия. – М. : Высшая школа, 1973. – 296 с.
86. **Добрынин, В. Ю.** Оценка тематического подобия текстовых документов / В. Ю. Добрынин, В. В. Клюев, И. С. Некрестьянов // Электронные библио-

теки: персективные методы и технологии, электронные коллекции : материалы Второй Всеросийской научной конференции. – Протвино, 2000. – С. 204–210.

87. **Блюменау, Д. И.** Информационный анализ/синтез для формирования вторичного потока документов / Д. И. Блюменау. – СПб. : Изд-во "Профессия", 2002. – 240 с. – (Серия "Специалист").

88. **Зубов, А. В.** Информационные технологии в лингвистике : учебное пособие для студентов лингвистических факультетов высших учебных заведений / А. В. Зубов, И. И. Зубова. – М. : Издательский центр "Академия", 2004. – 208 с.

89. **Зорич, В. А.** Математический анализ / В. А. Зорич. – М. : ФАЗИС, 1997.

90. **Акулич, И. Л.** Задачи нелинейного программирования // Математическое программирование в примерах и задачах / И. Л. Акулич. – М. : Высшая школа, 1986. – 319 с.

91. **Анфилатов, В. С.** Системный анализ в управлении : учебное пособие / В. С. Анфилатов, А. А. Емельянов, А. А. Кукушкин ; под ред. А. А. Емельянова. – М. : Финансы и статистика, 2002. – 368 с.: ил.

92. **Интрилигатор, М.** Математические методы оптимизации и экономическая теория: Пер. с англ. Г. И. Жуковой, Ф. Я. Кельмана / М. Интрилигатор. – М. : Айрис-пресс, 2002. – 576 с.: ил. – (Высшее образование).

93. **Сио, К. К.** Управленческая экономика : пер. с англ. / К. К. Сио. – М. : ИНФРА-М, 2000. – 671 с.

94. **Толкунов, А. А.** Алгоритмы формирования выборок типичных и полезных комментариев в ходе обработки результатов общественного обсуждения законопроектов / А. А. Толкунов // Информационные системы и технологии. – 2014. – № 1(81). – С. 47–58.

95. **Толкунов, А. А.** О погружении в массив комментариев с использованием типичных представителей мнений / А. А. Кукушкин, А. А. Толкунов // Труды Северо-Кавказского филиала Московского технического университета

связи и информатики. – Ростов-на-Дону. : ПЦ "Университет" СКФ МТУСИ, 2013. – 592 с. – С. 521–523.

96. **Бородащенко, А. Ю.** Алгоритм фильтрации текстовой информации на основе марковской модели / А. Ю. Бородащенко, В. А. Яковлев // Информационные технологии. – 2011. – № 5. – С. 2–5.

97. **Гусев, В. Д.** Тематический анализ и квазиредиректирование текста с использованием сканирующих статистик / В. Д. Гусев, Л. А. Мирошниченко, Н. В. Саломатина // Труды международной конференции Диалог-2005 "Компьютерная лингвистика и интеллектуальные технологии". – М. : Наука, 2005. – С. 121–125.

98. **Шеффе, Г.** Дисперсионный анализ / Г. Шеффе. – М. : Наука, 1980. – 512 с.

99. **Кобзарь, А. И.** Прикладная математическая статистика / А. И. Кобзарь. – М. : Физматлит, 2006.

100. **Боровиков, В. П.** STATISTICA. Искусство анализа данных на компьютере / В. П. Боровиков. – СПб. : Питер, 2003. – 688 с.

101. **Кацко, И. А.** Практикум по анализу данных на компьютере : учебно-практическое пособие / И. А. Кацко, Н. Б. Паклин. – М. : КолосС, 2009. – 278 с.

102. **Куценко, А. И.** Статистический анализ технологических процессов в среде Statistica и Excel : учебное пособие для вузов / А. И. Куценко [и др.]; Национальный исследовательский Томский политехнический университет (ТПУ), Новокузнецкий филиал (НФ). – Томск : Изд-во ТПУ, 2010. – 277 с.

103. **Васильев, А. Н.** Научные вычисления в Microsoft Excel / А. Н. Васильев. – М. : Диалектика, 2004. – 512 с.

104. **Информационные** системы : учебное пособие для студентов вузов по специальности 071900 – "Информационные системы в экономике" / под ред. В. Н. Волковой, Б. И. Кузина. – СПб. : Изд-во СПбГТУ, 1998. – 213 с.

105. **Толкунов, А. А.** Математические основы и программная архитектура системы оперативной аналитической обработки комментариев к обсуж-

даемым законопроектам / А. А. Толкунов // Системы управления и информационные технологии. – 2013. – № 3.1(53). – С. 173–177.

106. **Толкунов, А. А.** О построении системы оперативной аналитической обработки комментариев в целях повышения эффективности процесса анализа результатов общественного обсуждения законопроектов / А. А. Толкунов // Информационные технологии моделирования и управления. – 2013. – № 4(82). – С. 339–351.

107. **Васильев В. Г.** Методы автоматизированной обработки текстов / В. Г. Васильев, М. П. Кривенко. – М. : ИПИ РАН, 2008. – 305 с.

108. **Вальвачев, А. Н.** Программирование на языке Delphi : учебное пособие / А. Н. Вальвачев, К. А. Сурков, Д. А. Сурков, Ю. М. Четырько. – М., 2005.

109. **Классификатор** функциональных стилей текстов "StyleClassifier": свидетельство об официальной регистрации программы для ЭВМ № 2006610680 / А. А. Овсянников, А. А. Толкунов. – М. : ФГУ ФИПС, 2006.

110. **Браславский, П. И.** Автоматическая классификация документов Internet по стилям [Электронный ресурс] / П. И. Браславский. – Режим доступа: <http://www.sbras.ru/ws/el-pub-2000/29/>. – Дата обращения: 14.05.2013.

111. **Copeck, T.** Automating the Measurement of Linguistic Features to Help Classify Texts as Technical [Электронный ресурс] / T. Copeck, K. Barker, S. Delisle, S. Szpakowicz. – Режим доступа: https://oraprdnt.uqtr.quebec.ca/pls/public/docs/GSC21/F313_TALN_2K.pdf. – Дата обращения: 14.05.2013.

112. **Karlgreen, J.** Recognizing Text Genres with Simple Metrics Using Discriminant Analysis / J. Karlgreen, D. Cutting // Proceedings of COLING 94, Kyoto, Japan. – 2004. – Pp. 1071–1074.

113. **Michos, S. E.** Categorizing Texts by Using a Three-Level Functional Style Description [Электронный ресурс] / S. E. Michos, E. Stamatatos, N. Fakotakis, G. Kokkinakis. – Режим доступа: <http://www.icsd.aegean.gr/lecturers/stamatatos/papers/AIMSA96.pdf>. – Дата обращения: 14.05.2013.

114. **Система** анализа тональности публикаций СМИ "Тонус": свидетельство об официальной регистрации программы для ЭВМ № 2008614713 / А. А. Кукушкин, И. С. Полянский. – М. : ФГУ ФИПС, 2008.

115. **IBM** Cognos Business Intelligence V10.1 Handbook [Электронный ресурс]. – Режим доступа: <http://www.redbooks.ibm.com/redbooks/pdfs/sg247912.pdf>. – Дата обращения: 15.03.2013.

116. **Программа** семантической фильтрации текстов: свидетельство о государственной регистрации программы для ЭВМ № 2009612007 / А. Ю. Бородащенко, М. В. Бочков, А. Л. Салбиев. – М. : ФГУ ФИПС, 2009.

117. **Программа** атрибутивного анализа текстовых комментариев интернет-пользователей к проектам правовых актов "Comment Eliminator": свидетельство о государственной регистрации программы для ЭВМ № 2012616109 от 04.07.2012 г. / А. А. Кукушкин, А. А. Толкунов, Е. А. Шевченко. – М.: ФГУ ФИПС, 2012.

ПРИЛОЖЕНИЕ А

АНАЛИЗ СТРУКТУРЫ ГЛАВ И СТАТЕЙ ИССЛЕДУЕМЫХ
ЗАКОНОПРОЕКТОВ

Таблица А.1 – Сравнительный анализ структуры глав и статей законопроекта "О полиции" перед обсуждением и после обработки результатов обсуждения

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
1	1	1	3	3	–	–	3	3
	2	2	2	2	1 п. – 11 пп.	1 п. – 11 пп.	13	13
	3	3	1	2	–	–	1	2
	4	4	5	5	–	–	5	5
2	5	5	5	7	–	4 п. – 2 пп.	5	9
	6	6	4	5	–	–	4	5
	7	7	4	4	–	–	4	4
	8	8	8	6	–	–	8	6
	9	9	11	11	–	7 п. – 5 пп.	11	16
	10	10	7	6	–	–	7	6
	11	11	4	4	–	–	4	4
3	12	12	3	3	1 п. – 48 пп.	1 п. – 48 пп.	51	51
	13	13	3	3	1 п. – 45 пп.	1 п. – 45 пп.	48	48
4	14	14	7	16	2 п. – 14 пп.	2 п. – 14 пп.	21	30
	15	15	3	8	1 п. – 3 пп., 2 п. – 4 пп.	3 п. – 4 пп., 5 п. – 4 пп.	10	16
	16	16	4	5	1 п. – 5 пп.	2 п. – 5 пп.	9	10
	17	17	7	8	3 п. – 20 пп.	3 п. – 23 пп.	27	31
5	18	18	7	9	–	–	7	9
	19	19	10	9	–	–	10	9
	20	20	2	2	–	1 п. – 3 пп.	2	5
	21	21	3	3	1 п. – 11 пп., 2 п. – 13 пп.	1 п. – 11 пп., 2 п. – 14 пп.	27	28
	22	22	3	5	1 п. – 2 пп., 2 п. – 6 пп.	1 п. – 2 пп., 2 п. – 4 пп.	11	11
	23	23	7	6	1 п. – 7 пп., 3 п. – 5 пп.	1 п. – 7 пп., 3 п. – 4 пп.	19	17
	24	24	2	2	–	–	2	2
6	25	25	9	9	2 п. – 2 пп.	2 п. – 2 пп.	11	11
	26	26	2	4	1 п. – 5 пп.	1 п. – 5 пп.	7	9
	27	27	2	2	1 п. – 13 пп., 2 п. – 2 пп.	1 п. – 14 пп., 2 п. – 2 пп.	17	18

Продолжение таблицы А.1

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
6	28	28	6	3	1 п. – 17 пп., 3 п. – 6 пп.	1 п. – 17 пп., 3 п. – 5 пп.	29	25
	29	29	3	3	1 п. – 10 пп.	1 п. – 10 пп.	13	13
	30	–	5	–	1 п. – 19 пп., 4 п. – 2 пп.	–	26	–
	31	–	2	–	1 п. – 14 пп.	–	16	–
	32	30	9	7	–	–	9	7
	33	31	2	2	–	–	2	2
	34	32	1	1	–	–	1	1
	35	33	5	4	–	–	5	4
7	36	34	5	2	–	–	5	2
	37	35	7	8	3 п. – 5 пп.	3 п. – 4 пп.	12	12
	38	36	4	3	2 п. – 3 пп.	–	7	6
	39	39	2	2	–	–	2	2
	40	37	6	5	–	–	6	5
	41	38	4	6	–	–	4	6
	42	40	3	4	1 п. – 23 пп.	1 п. – 22 пп.	26	26
	43	41	3	3	–	–	3	3
8	44	42	2	2	–	–	2	2
	45	43	9	9	3 п. – 2 пп., 4 п. – 4 пп.	3 п. – 2 пп., 4 п. – 4 пп.	15	15
	46	44	3	3	–	–	3	3
	47	45	4	4	–	–	4	4
	48	46	3	6	2 п. – 5 пп.	6 п. – 5 пп.	8	11
9	49	47	4	4	–	–	4	4
	50	48	8	7	–	–	8	7
10	51	49	2	2	–	–	2	2
	52	50	4	4	–	–	4	4
	–	51	–	1	–	–	–	1
	53	52	2	1	–	–	2	1
	54	53	1	1	–	–	1	1
11	55	55	1	1	1 п. – 45 пп.	1 п. – 45 пп.	46	46
	56	54	10	11	–	–	10	11
	57	56	2	2	–	–	2	2

Таблица А.2 – Сравнительный анализ структуры глав и статей законопроекта "Об образовании" перед обсуждением и после обработки результатов обсуждения

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
1/1	1	1	1	2	–	–	1	2
	2	2	1	1	–	1 п. – 34 пп.	1	35
	3	3	5	2	2 п. – 14 пп.	1 п. – 12 пп.	19	14
	4	4	4	9	–	3 п. – 6 пп.	4	15
	5		11		1 п. – 5 пп.		16	
	6	–	3	–	–	–	3	–
	7	–	1	–	–	–	1	–
	8	5	12	5	–	5 п. – 3 пп.	12	8
	9	–	8	–	2 п. – 5 пп.	–	13	–
	10	–	8	–	–	–	8	–
	11	6	1	2	1 п. – 19 пп.	1 п. – 14 пп.	20	16
	12	7	10	10	1 п. – 4 пп., 3 п. – 2 пп., 6 п. – 4 пп., 7 п. – 7 пп., 8 п. – 5 пп.	1 п. – 4 пп., 3 п. – 2 пп., 6 п. – 4 пп., 7 п. – 7 пп., 8 п. – 5 пп.	32	32
	13	8	2	3	1 п. – 13 пп.	1 п. – 13 пп.	15	16
	14	9	3	4	1 п. – 6 пп.	1 п. – 7 пп.	9	11
2/2	15	10	3	7	1 п. – 5 пп., 3 п. – 4 пп.	1 п. – 5 пп., 4 п. – 4 пп., 5 п. – 4 пп.	12	20
	16	11	10	10	2 п. – 3 пп., 3 п. – 5 пп.	1 п. – 4 пп., 3 п. – 3 пп.	18	17
	17	–	4	–	–	–	4	–
	18	12	13	15	5 п. – 2 пп., 7 п. – 2 пп.	3 п. – 3 пп., 4 п. – 2 пп.	17	20
	(63)	13	12	11	–	–	12	11
	–	14	–	6	–	–	–	6
	(65)	15	7	3	3 п. – 2 пп.	3 п. – 5 пп.	9	8
	(66)	16	6	5	–	–	6	5
	19	17	6	5	–	1 п. – 2 пп.	6	7
	(67)	18	10	9		4 п. – 2 пп.		11
2/2	20	19	4	4	–	–	4	4
	21	(97)	4	6	–	–	4	6
	22	20	5	5	–	–	5	5
	23	–	7	–	–	–	7	–
3/3	24	21	4	2	1 п. – 2 пп.	–	6	2
	25	22	15	15	2 п. – 3 пп.	–	18	15
	26	23	8	6	2 п. – 4 пп., 3 п. – 2 пп.	2 п. – 4 пп., 3 п. – 2 пп., 4 п. – 6 пп.	14	18

Продолжение таблицы А.2

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
3/3	(114, 115)	24	9	6	—	—	9	6
	27	25	7	3	3 п. – 8 пп.	2 п. – 4 пп.	15	7
	28	26	6	6	—	6 п. – 2 пп.	6	8
	29	27	12	12	—	—	12	12
	30	29	4	3	2 п. – 5 пп.	2 п. – 6 пп.	9	9
	31	28	8	7	2 п. – 24 пп., 5 п. – 6 пп.	3 п. – 22 пп., 6 п. – 3 пп.	38	32
	—	30	—	4	—	—	—	4
	32	31	3	6	—	—	3	6
	—	32	—	5	—	—	—	5
4/4	33	33	2	4	1 п. – 7 пп.	1 п. – 9 пп.	9	13
	34	34	14	9	1 п. – 27 пп., 2 п. – 11 пп., 3 п. – 3 пп.	1 п. – 29 пп., 2 п. – 7 пп.,	55	45
	35	35	2	3	—	—	2	3
	36	41	6	6	—	1 п. – 10 пп., 4 п. – 4 пп.	6	20
	37	36	11	17	—	2 п. – 7 пп.	11	24
	38	37	5	4	—	—	5	4
	39	38	2	2	—	—	2	2
	40	39	3	4	—	—	3	4
	41	40	3	2	—	—	3	2
	—	42	—	6	—	2 п. – 4 пп.	—	10
	42	—	3	—	—	—	3	—
	43	43	3	12	—	1 п. – 5 пп.	3	17
	44		4		—		4	
	45	44	2	6	—	3 п. – 8 пп., 4 п. – 3 пп.	2	17
	46		3		1 п. – 6 пп.		9	
	47		3		—		3	
	—	45	—	6	—	1 п. – 3 пп.	—	9
5/5	48	46	6	2	—	—	6	2
	49	47	14	10	1 п. – 14 пп., 4 п. – 7 пп.	3 п. – 13 пп., 5 п. – 7 пп.	35	30
	50	48	6	4	1 п. – 10 пп.	1 п. – 11 пп.	16	15
	51	49	3	4	—	—	3	4
	—	50	—	3	—	2 п. – 4 пп., 3 п. – 2 пп.	—	9
	52	—	5	—	—	—	5	—
	53	51	9	14	1 п. – 6 пп.	1 п. – 4 пп.	15	18
	54	52	2	4	—	—	2	4
	55	—	8	—	—	—	8	—
6/6	56	53	1	4	1 п. – 8 пп.	—	9	4
	—	54	—	10	—	1 п. – 2 пп.	—	12
	57	—	8	—	5 п. – 4 пп.	—	12	—

Продолжение таблицы А.2

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
	58	55	8	9	—	—	8	9
	—	56	—	10	—	5 п. – 2 пп., 6 п. – 3 пп.	—	15
	59	—	5	—	2 п. – 6 пп.	—	11	—
	60	57	3	4	1 п. – 9 пп.	—	12	4
	61	61	10	5	1 п. – 2 пп., 2 п. – 3 пп., 3 п. – 2 пп.	1 п. – 2 пп., 2 п. – 3 пп.	17	10
	62	62	2	2	—	—	2	2
7/6	63	(13)	12	11	—	—	12	11
	64	—	3	—	—	—	3	—
	65	(15)	7	3	3 п. – 2 пп.	3 п. – 5 пп.	9	8
	66	(16)	6	5	—	—	6	5
	67	(18)	10	9	5 п. – 2 пп.	4 п. – 2 пп.	12	11
	68	58	5	11	—	—	5	11
		59		17		9 п. – 2 пп., 12 п. – 3 пп., 13 п. – 2 пп., 15 п. – 2 пп.		26
	69	60	13	16	1 п. – 2 пп. 11 п. – 2 пп.	1 п. – 2 пп., 6 п. – 2 пп., 7 п. – 5 пп., 10 п. – 2 пп.	17	27
	70	(106)	5	4	—	—	5	4
	71	(107)	5	15	5 п. – 2 пп.	5 п. – 2 пп., 15 п. – 2 пп.	7	19
8/12	—	89	—	5	—	2 п. – 8 пп.	—	13
	72	90	4	2	2 п. – 4 пп. 3 п. – 4 пп.	2 п. – 3 пп.	12	5
	73	91	28	16	9 п. – 5 пп., 14 п. – 4 пп., 18 п. – 3 пп., 19 п. – 2 пп., 20 п. – 2 пп.	5 п. – 2 пп., 12 п. – 3 пп., 15 п. – 4 пп.	44	25
	74	92	39	29	19 п. – 2 пп., 23 п. – 2 пп., 27 п. – 3 пп., 28 п. – 2 пп., 31 п. – 3 пп.	19 п. – 2 пп., 23 п. – 2 пп., 24 п. – 3 пп., 29 п. – 9 пп.	51	45
	75	93	6	10	4 п. – 2 пп.	5 п. – 2 пп.	8	12
	—	94	—	4	—	—	—	4
	—	95	—	6	—	—	—	6
	76	96	4	9	—	—	4	9
	77		2		—		2	
	(21)	97	4	6	—	—	4	6
	—	98	—	12	—	—	—	12

Продолжение таблицы А.2

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
9/13	78	—	2	—	1 п. – 3 пп.	—	5	—
	79	—	6	—	—	—	6	—
	80	99	4	5	—	—	4	5
	81	100	6	5	2 п. – 6 пп.	4 п. – 3 пп.	6	8
	82	101	10	3	3 п. – 3 пп., 8 п. – 2 пп.	—	15	8
	83	102	7	3	2 п. – 3 пп.	—	10	3
	84	103	3	6	—	—	3	6
	85	104	8	4	7 п. – 11 пп.	—	19	4
	86	—	4	—	—	—	4	—
	87	—	5	—	—	—	5	—
	88	—	6	—	3 п. – 3 пп.	—	9	—
10/7	89	64	7	3	—	—	7	3
	90	—	2	—	—	—	2	—
	91	65	2	7	—	—	2	7
	92	—	3	—	—	—	3	—
11/7	93	63	11	5	—	—	11	5
	—	66	—	12	—	—	—	12
	94	—	6	—	2 п. – 3 пп.	—	9	—
	95	67	3	6	—	—	3	6
	96	—	6	—	—	—	6	—
	97	—	6	—	—	—	6	—
	98	—	3	—	—	—	3	—
	99	—	5	—	—	—	5	—
	100	—	4	—	2 п. – 5 пп.	—	9	—
	101	—	3	—	—	—	3	—
12/9	102	73	6	9	—	—	6	9
	103	—	3	—	—	—	3	—
	104	—	3	—	—	—	3	—
	105	—	3	—	—	—	3	—
	106	74	4	3	—	—	4	3
13	107	68	6	6	1 п. – 2 пп.	—	8	6
	108	—	4	—	2 п. – 2 пп.	—	6	—
	109	—	3	—	—	—	3	—
	110	—	3	—	—	—	3	—
	111	—	2	—	—	—	2	—
14	112	69	7	8	—	8 п. – 4 пп.	7	12
	113	—	6	—	—	—	6	—
	114	(24)	6	6	—	—	6	6
	115		3		—		3	
	116	—	12	—	9 п. – 4 пп.	—	16	—
	117	70	13	10	—	—	13	10

Окончание таблицы А.2

Глава	Статья (номер)		Количество пунктов		Наличие подпунктов		Всего объектов комментирования	
	перед	после	перед	после	перед	после	перед	после
14	—	71	—	12	—	1 п. – 5 пп., 4 п. – 2 пп., 7 п. – 13 пп. 12 п. – 2 пп.	—	34
	118	—	4	—	—	—	4	—
	119	—	5	—	—	—	5	—
	120	72	7	2	2 п. – 6 пп.	2 п. – 5 пп.	13	7
15/10	121	75	4	5	—	—	4	5
	122	—	5	—	—	—	5	—
16/10	123	76	4	16	—	3 п. – 2 пп.	4	18
	124	—	7	—	—	—	7	—
	125	—	3	—	—	—	3	—
	126	—	2	—	—	—	2	—
17/11	127	77	7	5	—	—	7	5
	128	78	5	8	—	—	5	8
	—	79	—	12	—	—	—	12
	129	80	3	9	—	—	3	9
	130	81	11	10	11 п. – 6 пп.	1 п. – 5 пп., 10 п. – 4 пп.	17	19
	131	82	5	14	1 п. – 3 пп.	1 п. – 3 пп., 4 п. – 3 пп.	8	20
	132	83	12	21	12 п. – 4 пп.	2 п. – 4 пп.	16	25
	133	84	7	9	—	2 п. – 3 пп., 3 п. – 2 пп.	7	14
	—	85	—	7	—	1 п. – 3 пп.	—	10
	—	86	—	6	—	—	—	6
	134	87	7	12	—	—	7	12
	—	88	—	8	—	5 п. – 5 пп.	—	13
18/14	135	105	6	3	5 п. – 6 пп.	1 п. – 3 пп., 3 п. – 5 пп.	12	11
	136	—	4	—	—	—	4	—
	137	—	2	—	—	—	2	—
	(70)	106	5	4	—	—	5	4
	(71)	107	5	15	5 п. – 2 пп.	5 п. – 2 пп., 15 п. – 2 пп.	7	19
19	138	111	4	6	—	—	4	6
	139	108	7	13	1 п. – 5 пп., 2 п. – 4 пп.	1 п. – 8 пп., 2 п. – 16 пп., 5 п. – 7 пп., 13 п. – 2 пп.	16	49
	140	109	1	1	—	1 п. – 11 пп.	1	12
		110		1		1 п. – 104 пп.		105

Таблица А.3 – Статьи законопроекта "О полиции", отраженные по мощности (количество объектов комментирования)

Ранг	Статья	Мощность	Ранг	Статья	Мощность	Ранг	Статья	Мощность
1	12	51	20	15	10	39	11	4
2	13	48	21	19	10	40	41	4
3	55	46	22	56	10	41	47	4
4	28	29	23	16	9	42	49	4
5	17	27	24	32	9	43	52	4
6	21	27	25	8	8	44	1	3
7	30	26	26	48	8	45	43	3
8	42	26	27	50	8	46	46	3
9	14	21	28	10	7	47	20	2
10	23	19	29	18	7	48	24	2
11	27	17	30	26	7	49	33	2
12	31	16	31	38	7	50	39	2
13	45	15	32	40	6	51	44	2
14	2	13	33	4	5	52	51	2
15	29	13	34	5	5	53	53	2
16	37	12	35	35	5	54	57	2
17	9	11	36	36	5	55	3	1
18	22	11	37	6	4	56	34	1
19	25	11	38	7	4	57	54	1

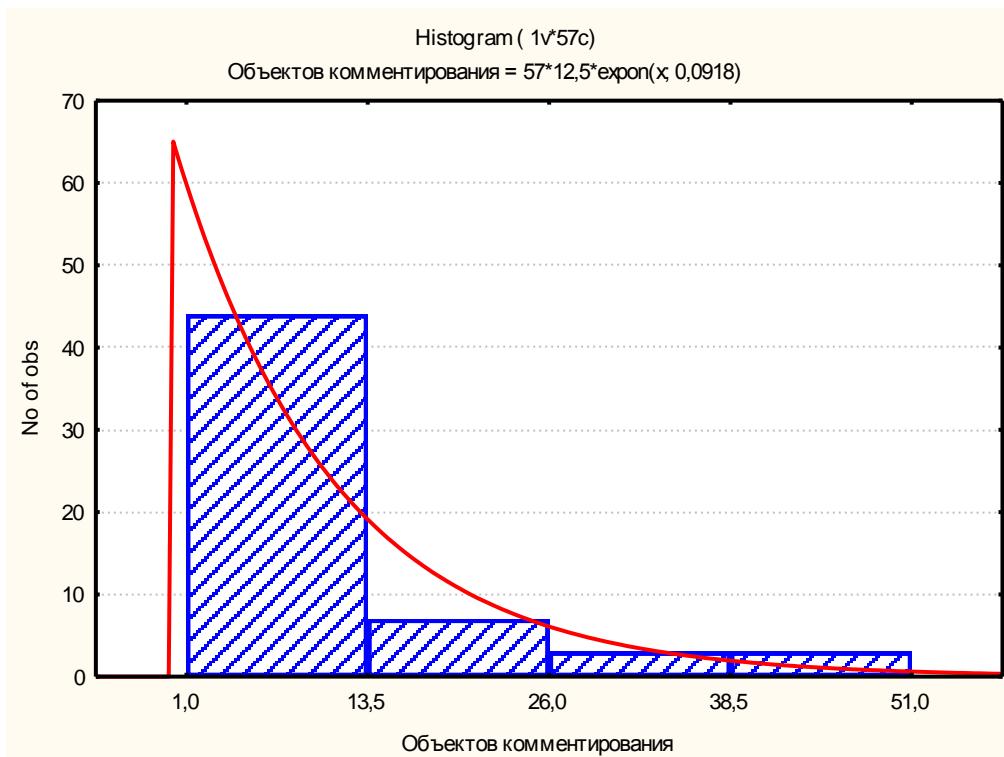


Рисунок А.1 – Распределение статей законопроекта "О полиции" по мощности

Таблица А.4 – Статьи законопроекта "Об образовании", отраженные по мощности

Ранг	Статья	Мощность	Ранг	Статья	Мощность	Ранг	Статья	Мощность
1	34	55	48	88	9	95	76	4
2	74	51	49	94	9	96	80	4
3	73	44	50	100	9	97	86	4
4	31	38	51	10	8	98	106	4
5	49	35	52	55	8	99	118	4
6	12	32	53	58	8	100	121	4
7	11	20	54	75	8	101	123	4
8	3	19	55	107	8	102	136	4
9	85	19	56	131	8	103	138	4
10	16	18	57	23	7	104	6	3
11	25	18	58	71	7	105	32	3
12	18	17	59	89	7	106	40	3
13	61	17	60	112	7	107	41	3
14	69	17	61	124	7	108	42	3
15	130	17	62	127	7	109	43	3
16	5	16	63	133	7	110	47	3
17	50	16	64	134	7	111	51	3
18	116	16	65	19	6	112	64	3
19	132	16	66	24	6	113	84	3
20	139	16	67	28	6	114	92	3
21	13	15	68	36	6	115	95	3
22	27	15	69	48	6	116	98	3
23	53	15	70	66	6	117	101	3
24	82	15	71	79	6	118	103	3
25	26	14	72	81	6	119	104	3
26	9	13	73	96	6	120	105	3
27	117	13	74	97	6	121	109	3
28	120	13	75	102	6	122	110	3
29	8	12	76	108	6	123	115	3
30	15	12	77	113	6	124	125	3
31	29	12	78	114	6	125	129	3
32	57	12	79	22	5	126	35	2
33	60	12	80	38	5	127	39	2
34	63	12	81	52	5	128	45	2
35	67	12	82	68	5	129	54	2
36	72	12	83	70	5	130	62	2
37	135	12	84	78	5	131	77	2
38	37	11	85	87	5	132	90	2
39	59	11	86	99	5	133	91	2
40	93	11	87	119	5	134	111	2
41	83	10	88	122	5	135	126	2
42	14	9	89	128	5	136	137	2
43	30	9	90	4	4	137	1	1
44	33	9	91	17	4	138	2	1
45	46	9	92	20	4	139	7	1
46	56	9	93	21	4	140	140	1
47	65	9	94	44	4			

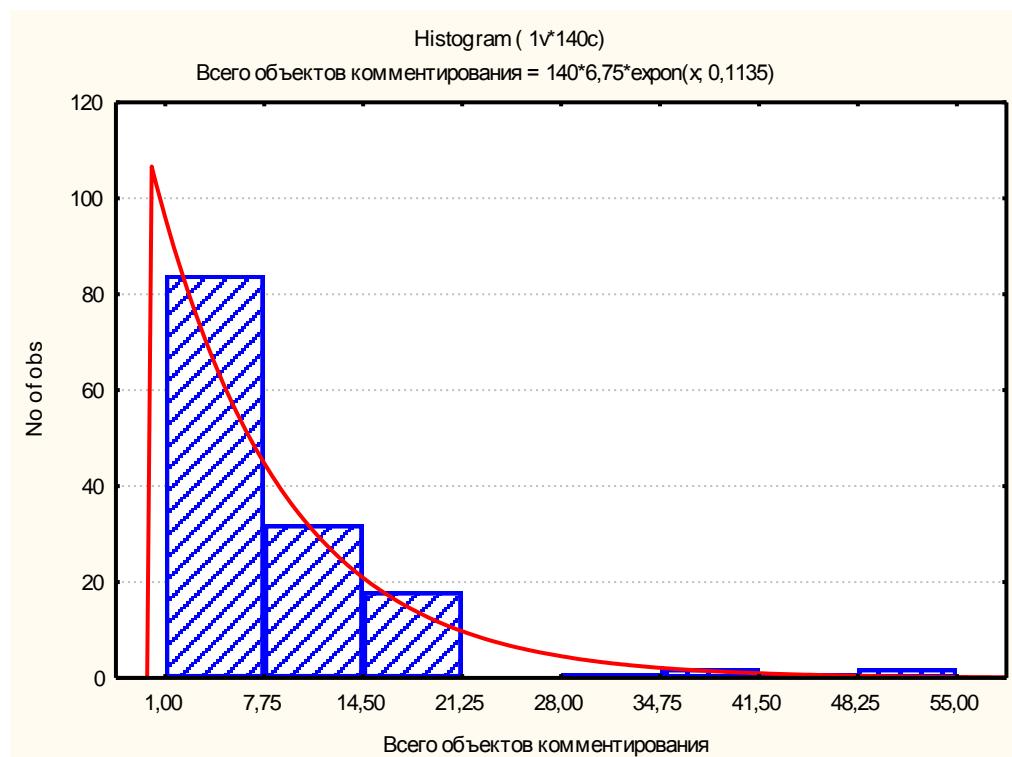


Рисунок А.2 – Распределение статей законопроекта "Об образовании" по мощности

ПРИЛОЖЕНИЕ Б

МНОЖЕСТВА И РАСПРЕДЕЛЕНИЯ ВЫСКАЗЫВАНИЙ В ОТОБРАННЫХ
ДЛЯ ИССЛЕДОВАНИЙ КЛАССАХ

Таблица Б.1 – Множество высказываний в комментариях класса "3-12-Од-В-Ми" к законопроекту "О полиции"

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "3-12-Од-В-Ми"$
1	Добавить в обязанности полиции розыск должников-граждан
2	Конкретизировать порядок приема заявлений и обращений граждан и организаций
3	Добавить в обязанности оказание содействия частным детективным агентствам
4	Добавить в обязанности полиции по обеспечению БДД ограничение эксплуатации неисправных дорог, информирование и контроль устранения несоответствий должностными лицами дорожных служб
5	Уточнить квалификацию действий по возбуждению уголовных дел, производству дознания, следственных действий как права либо обязанности полиции
6	Ввести временные нормативы на раскрытие преступлений, привлечение к суду виновных и организацию работы
7	Исключить обязанность по проведению индивидуальной работы с лицами, подготавливающими преступления, как противоречащую УК
8	Конкретизировать перечень пресекаемых административных правонарушений
9	Добавить исследования по материалам проверок сообщений о преступлениях
10	Уточнить порядок обеспечения сохранности найденных и сданных в полицию документов и имущества
11	Сообщать судам о начале и окончании назначенных ими наказаний
12	Добавить исполнение решений и поручений суда
13	Добавить обязанность контроля дорожных работ от экспертной оценки проекта до приемки выполненных работ
14	Добавить обязанность участия в лицензировании деятельности по проектированию, строительству, реконструкции дорог и дорожных сооружений
15	Добавить обязанность экспертной оценки программ подготовки водителей
16	Добавить обязанность участия в лицензировании образовательных учреждений, осуществляющих подготовку водителей
17	Добавить обязанность полиции защищать законные интересы и конституционные права, имущество граждан и организаций, блюсти честь и достоинство гражданина и человека, быть доброжелательной, тактичной, внимательной и вежливой
18	Добавить в обязанности недопустимость действий полиции, провоцирующих человека на совершение преступления
19	Устранить пересечение обязанностей с СК РФ по возбуждению "налоговых" дел
20	Согласовать формулировки пункта о допуске к работе в службе авиационной безопасности с требованиями пунктов ФЗ "О транспортной безопасности" и Воздушного Кодекса РФ
21	Конкретизировать механизм обеспечения явки граждан и должностных лиц для дачи объяснений
22	Добавить участие совместно со специализированными органами в сфере соблюдения земельного и природоохранного законодательства в мероприятиях по пресечению незаконных действий

Продолжение таблицы Б.1

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "3-12-Од-В-Ми"$
23	Заменить слова "содействие организаторам спортивных, зрелищных мероприятий..." на "охрана общественного порядка"
24	Удалить обязанность "Бюро находок" за исключением предметов, запрещенных к обращению или с ограниченным оборотом
25	Уточнить состав мероприятий по контролю соблюдения частными детективными и охранными организациями требований законодательства РФ
26	Добавить обязанность по согласованию маршрутов транспортных средств
27	Добавить обязанность проверять наличие у граждан навыков безопасного обращения с различными видами оружия, на приобретение которых выдана лицензия
28	Убрать обязанности по контролю соблюдения миграционного законодательства
29	Исключить право проведения экспертиз по уголовным и административным делам
30	Исключить охрану на договорной основе имущества граждан и организаций
31	Уточнить понятие "электронный документ"
32	УстраниТЬ обязанность розыска лиц, уклоняющихся от назначенных судом мер, как дублирование функций судебных приставов
33	Уточнить виды оружия, на приобретение которых полиция выдает лицензию
34	УстраниТЬ дублирование функций с ФСИН по охране и конвоированию заключенных
35	УстраниТЬ дублирование функций с Федеральным агентством воздушного транспорта по обеспечению авиационной безопасности и допуску к работе в службе авиационной безопасности
36	Узаконить понятие "доследственной проверки"
37	Заменить термин "имущество" перечислением всех защищаемых форм собственности
38	Исключить пункты обязанностей, не предполагающие ношения оружия и применения силы
39	Конкретизировать обязанности сотрудника полиции при обращении к любому лицу
40	Убрать обязанность согласования проведения спортивных, зрелищных и иных массовых мероприятий
41	Убрать обязанность "совместного" с организаторами собраний и митингов обеспечения общественного порядка
42	Дать определение понятию "общественное место"
43	Добавить ответственность за нарушение обязанностей или халатное бездействие
44	Конкретизировать формы содействия избирательным комиссиям и СМИ в ходе избирательных компаний
45	Добавить в подпункте 1 пункта 1 предлог "на" перед словом "жизнь"
46	Добавить в обязанности полиции предотвращение преступлений
47	Добавить обязанности по защите лиц на основе заявления об угрозе немедленной расправы
48	Добавить обязанности по защите лиц на основе заявления об утере документов или ограблении
49	Уточнить состав обязанностей следственного управления при ОВД
50	УстраниТЬ дублирование функций с органами опеки по профилактике безнадзорности несовершеннолетних и безответственного поведения родителей

Таблица Б.2 – Множество высказываний в комментариях класса "7-42-Н-В-Ми" к законопроекту "О полиции"

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "7-42-Н-В-Ми"$
1	Добавить в основания для увольнения обнаружение отклонений при прохождении психофизиологического исследования, а также положительного результата при тестированиях на алкогольную, наркотическую и иную токсическую зависимость
2	Уточнить основания для увольнения в связи с организационно-штатными мероприятиями
3	Убрать право выбора основания увольнения
4	Учесть различия между специальными и воинскими званиями
5	Добавить в основание "по выслуге срока службы, дающего право на пенсию" фразу "по инициативе сотрудника либо с его согласия"
6	Конкретизировать срок представления подложных документов при поступлении на службу по отношению к принимаемому закону
7	Указать предельный возраст пребывания на службе
8	Добавить в основания для увольнения по отрицательным пунктам слова "...без права восстановления"
9	Уточнить основание "в связи с восстановлением в должности сотрудника полиции..." фразой "... в случае добросовестного выполнения обязанностей дать возможность продолжить службу"
10	Урегулировать последовательность службы в полиции и армии
11	Добавить пункт "за взяточничество"
12	Ввести индивидуальные контракты для сотрудников полиции
13	Добавить основание "за совершение административного правонарушения"
14	Ввести ротацию начальников не реже раза в 3–5 лет
15	Заштитить подчиненных от произвола начальников
16	Убрать из основания "в связи с прекращением уголовного преследования по нереабилитирующем основаниям" фразу "за исключением уголовных дел частного обвинения"
17	Добавить основание "за совершение уголовного правонарушения"
18	Исключить практику увольнения "задним числом"
19	Конкретизировать проступки, порочащие честь сотрудника полиции
20	Исключить основание "по несоответствию замещаемой должности вследствие недостаточной квалификации по результатам аттестации"
21	Взять за основу основания для увольнения из трудового кодекса
22	Конкретизировать порядок увольнения с учетом льготной выслуги, работы на севере, выслуги более 20 лет
23	Конкретизировать срок службы, дающий право на пенсию
24	Конкретизировать грубые нарушения служебной дисциплины
25	Добавить наказание за использование штатных единиц не по назначению
26	Добавить ответственность за нарушения по окончании прохождения службы
27	Убрать срок давности из пункта "по нереабилитирующем основаниям"
28	Лишить сотрудника всех льгот при увольнении по отрицательным пунктам
29	Уточнить порядок увольнения при отчислении с очной и заочной форм обучения

Таблица Б.3 – Множество высказываний в комментариях класса "6-34-Н-Н-Ми" к законопроекту "О полиции"

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "6-34-Н-Н-Ми"$
1	Исключить привлечение к оперативно-розыскной деятельности сотрудников внутренней службы, юстиции и стажеров
2	Конкретизировать требования (статус, права и обязанности, уровень образования, степень подготовки, экипировка и пр.) к привлекаемым к охране общественного порядка и пр. задачам иных сотрудников органов внутренних дел (ОВД) и стажеров
3	Включить в перечень оснований привлечения сотрудников ОВД объявление режима ЧС, природные и техногенные катастрофы, ликвидацию последствий стихийных действий, проведение контртеррористических операций, проведение массовых мероприятий
4	Уточнить категории сотрудников ОВД, на которые распространяется действие статьи
5	Предоставить гражданину право удостовериться, что привлеченный сотрудник уполномочен совершать в отношении него действия
6	Убрать статью
7	Исключить привлечение потерпевших к процедуре дознания
8	Исключить процессуальные действия за счет заявителя

Таблица Б.4 – Множество высказываний в комментариях класса "4-34-ГП-Н-Ми" к законопроекту "Об образовании"

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "4-34-ГП-Н-Ми"$
1	Добавить право студентов на получение материальной помощи
2	Убрать право на поощрение за успехи в различных видах деятельности
3	Регламентировать законодательно перевод студентов
4	Добавить право студентов на получение социальной стипендии
5	Указать конкретно формы поощрения студентов
6	Не указывать возраст, в котором обучающиеся могут создавать общественные объединения
7	Применить термин "запрещается" в отношении принуждения обучающихся к вступлению и участию в деятельности общественно-политических организаций
8	Прописать меру ответственности принуждающих к вступлению и участию в деятельности общественно-политических организаций
9	Добавить обязанность образовательных учреждений создавать и регулярно пополнять библиотечно-информационные фонды в необходимом количестве
10	Добавить обязанность образовательных учреждений регулярно обновлять парк компьютерной техники
11	Разработать типовое положение о переводе студентов с платной формы обучения на бюджет
12	Дополнить норму о праве перевода на бюджет указанием на регламентацию порядка перевода уставом вуза и типовым положением
13	Регламентировать продолжительность каникул
14	Определить фонд материальной помощи в размере 25 % стипендиального фонда
15	Добавить право бесплатной публикации в изданиях образовательной организации

Продолжение таблицы Б.4

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "4-34-ГП-Н-Ми"$
16	Прописать размеры выплат материальной помощи и социальной стипендии
17	Прописать запрет на привлечение учащихся к отработке летней трудовой практики
18	Добавить право на получение ежегодного пособия в размере 2-месячной государственной стипендии аспиранта для приобретения научной литературы
19	Прописать поправку о том, что конкретный перечень работ, к которым допускается привлекать учащихся, должен быть перечислен в уставе школы
20	Указать условия и источники выплаты стипендии
21	Ввести нижний предел стипендии
22	Изменить стилистику первого абзаца с "обучающимся предоставляются права на" на "обучающиеся имеют право на"
23	Прописать порядок действий обучающихся по защите своих прав

Таблица Б.5 – Множество высказываний в комментариях класса "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

m_v	Содержание высказывания $sent_{v,m_v}$, $v = "1-11-ОД-Н-Ми"$
1	Добавить подготовку специалистов в педагогических вузах с целью воспитания и обучения одаренных детей
2	Ввести десятибалльную систему оценки знаний учащихся
3	Ввести мониторинг качества образования
4	Ввести общественный контроль образования
5	Ввести прозрачные информационные процедуры в образовании
6	Добавить установление общих принципов и нормативов финансового обеспечения образовательной деятельности
7	Ввести обеспечение учебниками и учебными пособиями за счет средств федерального бюджета
8	Добавить разработку и реализацию государственной политики в области охраны здоровья и гармоничного развития обучающихся
9	Добавить установление перечня олимпиад школьников и порядка их проведения
10	Добавить установление норм труда, нормативов и порядка его оплаты в федеральных государственных образовательных учреждениях
11	Добавить установление порядка разработки и использования дистанционных образовательных технологий
12	Добавить установление порядка предоставления гражданам и погашения ими государственного образовательного кредита
13	Добавить разработку и утверждение типовых положений об образовательных учреждениях
14	Добавить установление форм и порядка проведения государственной (итоговой) аттестации обучающихся
15	Добавить установление мер социальной поддержки различных категорий обучающихся, педагогических, руководящих и иных работников образовательных организаций
16	Добавить установление образовательного ценза и особенностей регулирования труда педагогических работников

Продолжение таблицы Б.5

m_v	Содержание высказывания $sent_{v, m_v}$, $v = "1-11-ОД-Н-Ми"$
17	Добавить утверждение квалификационных характеристик должностей, профессиональных квалификационных групп и квалификационных уровней профессиональных квалификационных групп работников образования
18	Добавить утверждение базовых окладов (базовых должностных окладов), базовых ставок заработной платы работников по профессиональным квалификационным группам работников государственных и муниципальных образовательных учреждений
19	Добавить установление порядка аттестации педагогических работников и работников, занимающих должности научно-педагогических работников
20	Добавить установление механизмов создания, утверждения, применения и пересмотра федеральных государственных образовательных стандартов
21	Добавить упоминание негосударственных образовательных учреждений в пункт о финансировании выполнения государственного задания
22	Добавить установление периодически обновляемых базовых штатных расписаний
23	Добавить обеспечение защиты федеральных баз данных в сфере образования

Таблица Б.6 – Множество высказываний в комментариях класса "1-1-ОД-Н-Ми" к законопроекту "Об образовании"

m_v	Содержание высказывания $sent_{v, m_v}$, $v = "1-1-ОД-Н-Ми"$
1	Добавить пункт о приоритетном положении образования в государстве и обществе по отношению ко всем сферам деятельности
2	Конкретизировать термин "государственные гарантии прав и свобод в образовании Российской Федерации"
3	Добавить обеспечение социальной защищенности обучаемых с ограниченными возможностями здоровья
4	Добавить ответственность органов власти за финансирование программ дополнительного образования
5	Добавить ответственность органов власти за нарушения в сфере образования
6	Добавить указание на наличие обязанностей у участников отношений в сфере образования
7	Исключить термин "начала правового регулирования"
8	Разделить закон на отдельные кодексы или иные акты для каждой ступени образования
9	Изложить акт в форме единого кодекса Российской Федерации об образовании
10	Добавить установление принципов воспитания
11	Указать вместо "интересов государства" – "интересов человека, семьи и общества"
12	Заменить "гарантии прав и свобод в образовании" на "гарантии прав и свобод участников образовательного процесса"
13	Изложить положения статьи отдельными пунктами, а не сплошным текстом
14	Заменить термин "отрасль права" на "отрасль законодательства"
15	Заменить "гарантии прав и свобод в образовании" на "гарантии прав и свобод в системе образования"

Таблица Б.7 – Распределение высказываний по комментариям класса "7-42-Н-В-Ми" к законопроекту "О полиции"

Таблица Б.8 – Распределение высказываний по комментариям класса "3-12-Од-В-Ми" законопроекта "О полиции"

Продолжение таблицы Б.8

Таблица Б.9 – Распределение высказываний по комментариям класса "6-34-Н-Н-Ми" к законопроекту "О полиции"

Ид. КОММ.	m_V							
	1	2	3	4	5	6	7	8
2393	+	+						
2505		+	+					
5601				+				
6205		+			+			
6239	+						+	
14322	+							
15793		+						
15907							+	+
17329				+				

Таблица Б.10 – Распределение высказываний по комментариям класса "4-34-ГП-Н-Ми" к законопроекту "Об образовании"

Продолжение таблицы Б.10

Таблица Б.11 – Распределение высказываний по комментариям класса "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

Таблица Б.12 – Распределение высказываний по комментариям класса "1-1-ОД-Н-Ми" к законопроекту "Об образовании"

ПРИЛОЖЕНИЕ В

РАСЧЕТ СРЕДНЕГО ТЕМПА ПРИРОСТА ПОЛНОТЫ ОТБОРА ВЫСКАЗЫВАНИЙ В ВЫБОРКЕ ТИПИЧНЫХ КОММЕНТАРИЕВ

Таблица В.1 – Расчет пошагового накопления полноты отбора высказываний для исходного множества комментариев класса "7-42-Н-В-Ми" к законопроекту "О полиции"

Таблица В.2 – Расчет пошагового накопления полноты отбора высказываний для отранжированного по типичности множества комментариев класса "7-42-Н-В-Ми" к законопроекту "О полиции"

Число класте- ров	Ид. комм.	m_v																													Кол. выск.	Накоплено высказываний				Полнота				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		б/вес, б/повт	вес, б/повт	б/вес, повт	вес, повт	R1	R2	R3	R4	
1	2839	1	1				1	1	1	1	1																				7	7	15	7	15	0,175	0,202703	0,241379	0,375	
2	1779				1																											1	8	16	8	16	0,2	0,216216	0,275862	0,4
3	10872																	1														1	9	17	9	17	0,225	0,22973	0,310345	0,425
4	15920																														1	10	18	10	18	0,25	0,243243	0,344828	0,45	
5	13869																		1													1	11	19	11	19	0,275	0,256757	0,37931	0,475
6	13331																		1													1	12	20	12	20	0,3	0,27027	0,413793	0,5
7	15430																		1													1	13	22	12	20	0,325	0,297297	0,413793	0,5
8	16544	1																													1	14	26	12	20	0,35	0,351351	0,413793	0,5	
9	15105	1																													1	15	30	12	20	0,375	0,405405	0,413793	0,5	
10	16593																			1												1	16	33	13	23	0,4	0,445946	0,448276	0,575
11	11828																		1												1	17	34	14	24	0,425	0,459459	0,482759	0,6	
12	1658		1																											1	18	35	15	25	0,45	0,472973	0,517241	0,625		
13	16736																				1									1	19	36	16	26	0,475	0,486486	0,551724	0,65		
14	18525																				1									1	20	37	17	27	0,5	0,5	0,586207	0,675		
15	19074																				1									1	21	38	18	28	0,525	0,513514	0,62069	0,7		
16	7109																		1										1	22	39	19	29	0,55	0,527027	0,655172	0,725			
17	14937																			1	1								2	24	43	20	30	0,6	0,581081	0,689655	0,75			
18	3875			1																1									1	25	46	20	30	0,625	0,621622	0,689655	0,75			
19	12232																			1									1	26	48	21	32	0,65	0,648649	0,724138	0,8			
20	2715				1																								1	27	49	22	33	0,675	0,662162	0,758621	0,825			
21	5927																	1										1	28	50	23	34	0,7	0,675676	0,793103	0,85				
22	13825	1																		1									2	30	55	23	34	0,75	0,743243	0,793103	0,85			
23	16281																				1	1							2	32	57	25	36	0,8	0,77027	0,862069	0,9			
24	5639																	1										1	33	58	26	37	0,825	0,783784	0,896552	0,925				
25	16327																		1									1	34	61	26	37	0,85	0,824324	0,896552	0,925				
26	10327																	1										1	35	62	27	38	0,875	0,837838	0,931034	0,95				
27	359	1	1	1															1									3	38	70	28	39	0,95	0,945946	0,965517	0,975				
28	7041																		1									2	40	74	29	40	1	1	1	1				
Вес высказыв.		1	4	3	1	1	1	3	1	1	2	1	1	1	1	1	1	1	1	2	1	1	1	3	1	1	1	1	1	40										

Таблица В.3 – Расчет пошагового накопления полноты отбора высказываний для исходного множества комментариев класса "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

Число кластеров	Ид. комм.	m_v																							Кол. высск.	Накоплено высказываний				Полнота			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		б/вес, б/повт	вес, б/повт	б/вес, повт	вес, повт	$R1_{исх}$	$R2_{исх}$	$R3_{исх}$	$R4_{исх}$
6	190	1	1																						2	2	2	2	0,054054	0,020619	0,086957	0,054054	
4	468			1																					1	3	4	3	4	0,081081	0,041237	0,130435	0,108108
3	510				1	1																			2	5	6	5	6	0,135135	0,061856	0,217391	0,162162
15	513						1																		1	6	11	6	11	0,162162	0,113402	0,26087	0,297297
11	1052							1																	1	7	12	7	12	0,189189	0,123711	0,304348	0,324324
16	1872								1																1	8	14	8	14	0,216216	0,14433	0,347826	0,378378
17	2343									1	1	1	1											4	12	24	12	24	0,324324	0,247423	0,521739	0,648649	
14	2346																								2	14	26	14	26	0,378378	0,268041	0,608696	0,702703
1	2823										1														6	20	39	19	33	0,540541	0,402062	0,826087	0,891892
12	2956										1	1	1											4	24	51	19	33	0,648649	0,525773	0,826087	0,891892	
8	3615										1													2	26	59	19	33	0,702703	0,608247	0,826087	0,891892	
13	3616											1												1	27	64	19	33	0,72973	0,659794	0,826087	0,891892	
19	3925											1												1	28	69	19	33	0,756757	0,71134	0,826087	0,891892	
18	4006		1																					1	29	71	19	33	0,783784	0,731959	0,826087	0,891892	
7	4129																							1	30	72	20	34	0,810811	0,742268	0,869565	0,918919	
10	5177												1											2	32	83	20	34	0,864865	0,85567	0,869565	0,918919	
2	7077																							1	33	84	21	35	0,891892	0,865979	0,913043	0,945946	
9	7754										1												1	3	36	96	22	0,972973	0,989691	0,956522	0,972973		
5	8095																							1	1	37	97	23	0,972973	0,989691	0,956522	0,972973	
Вес высказыв.		1	1	2	1	1	5	1	2	1	6	1	2	1	1	2	1	1	2	1	1	1	1	37									

Таблица В.4 – Расчет пошагового накопления полноты отбора высказываний для отранжированного по типичности множества комментариев класса "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

Число кластеров	Ид. комм.	m_v																							Кол. высск.	Накоплено высказываний				Полнота			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		б/вес, б/повт	вес, б/повт	б/вес, повт	вес, повт	R1	R2	R3	R4
1	2823									1				1	1	1	1	1	1					6	6	13	6	13	0,162162	0,134021	0,26087	0,351351	
2	7077																								1	7	14	7	14	0,189189	0,14433	0,304348	0,378378
3	510			1	1																			2	9	16	9	16	0,243243	0,164948	0,391304	0,432432	
4	468		1																					1	10	18	10	18	0,27027	0,185567	0,434783	0,486486	
5	8095																							1	1	11	19	11	19	0,297297	0,195876	0,478261	0,513514
6	190	1	1																					2	13	21	13	21	0,351351	0,216495	0,565217	0,567568	
7	4129																		1					1	14	22	14	22	0,378378	0,226804	0,608696	0,594595	
8	3615									1				1										2	16	30	14	22	0,432432	0,309278	0,608696	0,594595	
9	7754				1					1														1	3	19	42	16	28	0,513514	0,43299	0,695652	0,756757
10	5177				1					1														2	21	53	16	28	0,567568	0,546392	0,695652	0,756757	
11	1052					1																		1	22	54	17	29	0,594595	0,556701	0,73913	0,783784	
12	2956						1		1	1									1				4	26	66	19	33	0,702703	0,680412	0,826087	0,891892		
13	3616					1																		1	27	71	19	33	0,72973	0,731959	0,826087	0,891892	
14	2346													1	1									2	29	73	21	35	0,783784	0,752577	0,913043	0,945946	
15	513				1																			1	30	78	21	35	0,810811	0,804124	0,913043	0,945946	
16	1872						1																	1	31	80	21	35	0,837838	0,824742	0,913043	0,945946	
17	2343								1	1	1	1											4	35	90	23	37	0,945946	0,927835	1	1		
18	4006		1																					1	36	92	23	37	0,972973	0,948454	1	1	
19	3925						1																1	37	97	23	37	1	1	1	1		
Вес высказыв.		1	1	2	1	1	5	1	2	1	6	1	2	1	1	2	1	1	2	1	1	1	1	37									

Таблица В.5 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "3-12-Од-В-Ми" к законопроекту "О полиции"

Квота для класса	Бюджет времени, Td/Td_{\max}																				
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1	
3-12-Од-В-Ми	3	6	9	12	15	17	20	23	26	29	32	35	38	41	44	46	49	52	55	58	
Полнота отбора высказываний																					
R1	0,042857	0,085714	0,171429	0,214286	0,257143	0,285714	0,385714	0,457143	0,5	0,542857	0,585714	0,628571	0,671429	0,728571	0,785714	0,814286	0,857143	0,914286	0,957143	1	
R2	0,039683	0,079365	0,142857	0,18254	0,253968	0,277778	0,349206	0,404762	0,468254	0,507937	0,539683	0,579365	0,619048	0,68254	0,746032	0,785714	0,849206	0,920635	0,968254	1	
R3	0,06	0,12	0,24	0,3	0,34	0,36	0,5	0,58	0,6	0,64	0,68	0,72	0,78	0,84	0,86	0,86	0,88	0,92	0,96	1	
R4	0,071429	0,142857	0,257143	0,328571	0,428571	0,442857	0,571429	0,642857	0,657143	0,685714	0,714286	0,757143	0,828571	0,871429	0,885714	0,885714	0,9	0,928571	0,971429	1	
R1 _{исх}	0,042857	0,085714	0,128571	0,171429	0,285714	0,314286	0,357143	0,4	0,442857	0,5	0,571429	0,657143	0,7	0,742857	0,785714	0,814286	0,857143	0,9	0,942857	1	
R2 _{исх}	0,039683	0,071429	0,103175	0,134921	0,253968	0,301587	0,333333	0,373016	0,404762	0,468254	0,539683	0,650794	0,68254	0,722222	0,753968	0,809524	0,873016	0,920635	0,968254	1	
R3 _{исх}	0,06	0,12	0,18	0,24	0,38	0,4	0,46	0,5	0,54	0,6	0,66	0,7	0,76	0,8	0,84	0,84	0,88	0,9	0,92	1	
R4 _{исх}	0,071429	0,128571	0,185714	0,242857	0,428571	0,442857	0,5	0,542857	0,571429	0,657143	0,714286	0,742857	0,8	0,842857	0,871429	0,871429	0,914286	0,928571	0,942857	1	
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																					
$\Delta R1$	4,3%	4,3%	8,6%	4,3%	4,3%	2,9%	10,0%	7,1%	4,3%	4,3%	4,3%	4,3%	4,3%	4,3%	5,7%	5,7%	2,9%	4,3%	5,7%	4,3%	4,3%
$\Delta R2$	4,0%	4,0%	6,3%	4,0%	7,1%	2,4%	7,1%	5,6%	6,3%	4,0%	3,2%	4,0%	4,0%	4,0%	6,3%	6,3%	4,0%	6,3%	7,1%	4,8%	3,2%
$\Delta R3$	6,0%	6,0%	12,0%	6,0%	4,0%	2,0%	14,0%	8,0%	2,0%	4,0%	4,0%	4,0%	6,0%	6,0%	2,0%	0,0%	2,0%	4,0%	4,0%	4,0%	4,0%
$\Delta R4$	7,1%	7,1%	11,4%	7,1%	10,0%	1,4%	12,9%	7,1%	1,4%	2,9%	2,9%	4,3%	7,1%	4,3%	1,4%	0,0%	1,4%	2,9%	4,3%	2,9%	2,9%
$\Delta R1_{исх}$	4,3%	4,3%	4,3%	4,3%	11,4%	2,9%	4,3%	4,3%	4,3%	5,7%	7,1%	8,6%	4,3%	4,3%	2,9%	4,3%	4,3%	4,3%	4,3%	4,3%	5,7%
$\Delta R2_{исх}$	4,0%	3,2%	3,2%	3,2%	11,9%	4,8%	3,2%	4,0%	3,2%	6,3%	7,1%	11,1%	3,2%	4,0%	3,2%	5,6%	6,3%	4,8%	4,8%	3,2%	8,0%
$\Delta R3_{исх}$	6,0%	6,0%	6,0%	6,0%	14,0%	2,0%	6,0%	4,0%	4,0%	6,0%	6,0%	4,0%	6,0%	4,0%	0,0%	4,0%	2,0%	2,0%	2,0%	2,0%	8,0%
$\Delta R4_{исх}$	7,1%	5,7%	5,7%	5,7%	18,6%	1,4%	5,7%	4,3%	2,9%	8,6%	5,7%	2,9%	5,7%	4,3%	2,9%	0,0%	4,3%	1,4%	1,4%	5,7%	5,7%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																					
$(\Delta R1)_{cp.}$	4,3%	4,3%	5,7%	5,4%	5,1%	4,8%	5,5%	5,7%	5,6%	5,4%	5,3%	5,2%	5,2%	5,2%	5,2%	5,1%	5,0%	5,1%	5,0%	5,0%	5,0%
$(\Delta R2)_{cp.}$	4,0%	4,0%	4,8%	4,6%	5,1%	4,6%	5,0%	5,1%	5,2%	5,1%	4,9%	4,8%	4,8%	4,9%	5,0%	4,9%	5,0%	5,1%	5,1%	5,0%	5,0%
$(\Delta R3)_{cp.}$	6,0%	6,0%	8,0%	7,5%	6,8%	6,0%	7,1%	7,3%	6,7%	6,4%	6,2%	6,0%	6,0%	5,7%	5,4%	5,2%	5,1%	5,0%	5,1%	5,0%	5,0%
$(\Delta R4)_{cp.}$	7,1%	7,1%	8,6%	8,2%	8,6%	7,4%	8,2%	8,0%	7,3%	6,9%	6,5%	6,3%	6,4%	6,2%	5,9%	5,5%	5,3%	5,2%	5,1%	5,0%	5,0%
$(\Delta R1_{исх})_{cp.}$	4,3%	4,3%	4,3%	4,3%	5,7%	5,2%	5,1%	5,0%	4,9%	5,0%	5,2%	5,5%	5,4%	5,3%	5,2%	5,1%	5,0%	5,0%	5,0%	5,0%	5,0%
$(\Delta R2_{исх})_{cp.}$	4,0%	3,6%	3,4%	3,4%	5,1%	5,0%	4,8%	4,7%	4,5%	4,7%	4,9%	5,4%	5,3%	5,2%	5,0%	5,1%	5,1%	5,1%	5,1%	5,0%	5,0%
$(\Delta R3_{исх})_{cp.}$	6,0%	6,0%	6,0%	6,0%	7,6%	6,7%	6,6%	6,3%	6,0%	6,0%	5,8%	5,8%	5,7%	5,6%	5,3%	5,2%	5,0%	4,8%	5,0%	5,0%	5,0%
$(\Delta R4_{исх})_{cp.}$	7,1%	6,4%	6,2%	6,1%	8,6%	7,4%	7,1%	6,8%	6,3%	6,6%	6,5%	6,2%	6,0%	5,8%	5,4%	5,2%	5,0%	5,0%	5,0%	5,0%	5,0%
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																					
$\Delta R1_{cp.}$	0,0%	0,0%	1,4%	1,1%	-0,6%	-0,5%	0,4%	0,7%	0,6%	0,4%	0,1%	-0,2%	-0,2%	-0,1%	0,0%	0,0%	0,0%	0,1%	0,1%	0,0%	0,0%
$\Delta R2_{cp.}$	0,0%	0,4%	1,3%	1,2%	0,0%	-0,4%	0,2%	0,4%	0,7%	0,4%	0,0%	-0,6%	-0,5%	-0,3%	-0,1%	-0,1%	-0,1%	0,0%	0,0%	0,0%	0,0%
$\Delta R3_{cp.}$	0,0%	0,0%	2,0%	1,5%	-0,8%	-0,7%	0,6%	1,0%	0,7%	0,4%	0,2%	0,2%	0,2%	0,3%	0,1%	0,1%	0,0%	0,1%	0,2%	0,0%	0,0%
$\Delta R4_{cp.}$	0,0%	0,7%	2,4%	2,1%	0,0%	0,0%	1,0%	1,3%	1,0%	0,3%	0,0%	0,1%	0,2%	0,2%	0,1%	0,1%	-0,1%	0,0%	0,2%	0,0%	0,0%

Таблица В.6 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "7-42-Н-В-Ми" к законопроекту "О полиции"

Квота для класса	Бюджет времени, Td/Td_{\max}																			
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
7-42-Н-В-Ми	1	3	4	6	7	8	10	11	13	14	15	17	18	20	21	22	24	25	27	28
Полнота отбора высказываний																				
R1	0,175	0,225	0,25	0,3	0,325	0,35	0,4	0,425	0,475	0,5	0,525	0,6	0,625	0,675	0,7	0,75	0,825	0,85	0,95	1
R2	0,202703	0,22973	0,243243	0,27027	0,297297	0,351351	0,445946	0,459459	0,486486	0,5	0,513514	0,581081	0,621622	0,662162	0,675676	0,743243	0,783784	0,824324	0,945946	1
R3	0,241379	0,310345	0,344828	0,413793	0,413793	0,413793	0,448276	0,482759	0,551724	0,586207	0,62069	0,689655	0,689655	0,758621	0,793103	0,793103	0,896552	0,896552	0,965517	1
R4	0,375	0,425	0,45	0,5	0,5	0,5	0,575	0,6	0,65	0,675	0,7	0,75	0,75	0,825	0,85	0,85	0,925	0,925	0,975	1
R1 _{исх}	0,075	0,125	0,15	0,35	0,375	0,4	0,475	0,5	0,55	0,575	0,6	0,675	0,725	0,775	0,8	0,85	0,9	0,925	0,975	1
R2 _{исх}	0,108108	0,135135	0,148649	0,391892	0,405405	0,418919	0,486486	0,5	0,527027	0,554054	0,567568	0,648649	0,702703	0,783784	0,797297	0,824324	0,918919	0,959459	0,986486	1
R3 _{исх}	0,103448	0,172414	0,206897	0,37931	0,413793	0,448276	0,517241	0,551724	0,62069	0,655172	0,689655	0,724138	0,793103	0,793103	0,827586	0,896552	0,896552	0,965517	1	
R4 _{исх}	0,2	0,25	0,275	0,475	0,5	0,525	0,575	0,6	0,65	0,7	0,725	0,75	0,85	0,85	0,875	0,925	0,925	0,925	0,975	1
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1$	17,5%	5,0%	2,5%	5,0%	2,5%	2,5%	5,0%	2,5%	5,0%	2,5%	2,5%	7,5%	2,5%	5,0%	2,5%	5,0%	7,5%	2,5%	10,0%	5,0%
$\Delta R2$	20,3%	2,7%	1,4%	2,7%	2,7%	5,4%	9,5%	1,4%	2,7%	1,4%	1,4%	6,8%	4,1%	4,1%	1,4%	6,8%	4,1%	4,1%	12,2%	5,4%
$\Delta R3$	24,1%	6,9%	3,4%	6,9%	0,0%	0,0%	3,4%	3,4%	6,9%	3,4%	3,4%	6,9%	0,0%	6,9%	3,4%	0,0%	10,3%	0,0%	6,9%	3,4%
$\Delta R4$	37,5%	5,0%	2,5%	5,0%	0,0%	0,0%	7,5%	2,5%	5,0%	2,5%	2,5%	5,0%	0,0%	7,5%	2,5%	0,0%	7,5%	0,0%	5,0%	2,5%
$\Delta R1_{исх}$	7,5%	5,0%	2,5%	20,0%	2,5%	2,5%	7,5%	2,5%	5,0%	2,5%	2,5%	7,5%	5,0%	5,0%	2,5%	5,0%	5,0%	2,5%	5,0%	2,5%
$\Delta R2_{исх}$	10,8%	2,7%	1,4%	24,3%	1,4%	1,4%	6,8%	1,4%	2,7%	2,7%	1,4%	8,1%	5,4%	8,1%	1,4%	2,7%	9,5%	4,1%	2,7%	1,4%
$\Delta R3_{исх}$	10,3%	6,9%	3,4%	17,2%	3,4%	3,4%	6,9%	3,4%	3,4%	3,4%	3,4%	6,9%	0,0%	3,4%	6,9%	0,0%	0,0%	6,9%	3,4%	
$\Delta R4_{исх}$	20,0%	5,0%	2,5%	20,0%	2,5%	2,5%	5,0%	2,5%	5,0%	2,5%	2,5%	10,0%	0,0%	2,5%	5,0%	0,0%	0,0%	5,0%	2,5%	
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$(\Delta R1)_{cp}$	17,5%	11,3%	8,3%	7,5%	6,5%	5,8%	5,7%	5,3%	5,3%	5,0%	4,8%	5,0%	4,8%	4,8%	4,7%	4,7%	4,9%	4,7%	5,0%	5,0%
$(\Delta R2)_{cp}$	20,3%	11,5%	8,1%	6,8%	5,9%	5,9%	6,4%	5,7%	5,4%	5,0%	4,7%	4,8%	4,8%	4,8%	4,7%	4,5%	4,6%	4,6%	5,0%	5,0%
$(\Delta R3)_{cp}$	24,1%	15,5%	11,5%	10,3%	8,3%	6,9%	6,4%	6,0%	6,1%	5,9%	5,6%	5,7%	5,3%	5,4%	5,3%	5,0%	5,3%	5,0%	5,1%	5,0%
$(\Delta R4)_{cp}$	37,5%	21,3%	15,0%	12,5%	10,0%	8,3%	8,2%	7,5%	7,2%	6,8%	6,4%	6,3%	5,8%	5,9%	5,7%	5,3%	5,4%	5,1%	5,0%	
$(\Delta R1_{исх})_{cp}$	7,5%	6,3%	5,0%	8,8%	7,5%	6,7%	6,8%	6,3%	6,1%	5,8%	5,5%	5,6%	5,5%	5,3%	5,3%	5,1%	5,1%	5,0%		
$(\Delta R2_{исх})_{cp}$	10,8%	6,8%	5,0%	9,8%	8,1%	7,0%	6,9%	6,3%	5,9%	5,5%	5,2%	5,4%	5,4%	5,6%	5,3%	5,2%	5,4%	5,3%	5,2%	5,0%
$(\Delta R3_{исх})_{cp}$	10,3%	8,6%	6,9%	9,5%	8,3%	7,5%	7,4%	6,9%	6,9%	6,6%	6,3%	6,0%	6,1%	5,7%	5,5%	5,6%	5,3%	5,0%	5,1%	5,0%
$(\Delta R4_{исх})_{cp}$	20,0%	12,5%	9,2%	11,9%	10,0%	8,8%	8,2%	7,5%	7,2%	7,0%	6,6%	6,3%	6,5%	6,1%	5,8%	5,8%	5,4%	5,1%	5,0%	
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1_{cp}$	10,0%	5,0%	3,3%	-1,3%	-1,0%	-0,8%	-1,1%	-0,9%	-0,8%	-0,7%	-0,7%	-0,6%	-0,8%	-0,7%	-0,7%	-0,6%	-0,4%	-0,4%	-0,1%	0,0%
$\Delta R2_{cp}$	9,5%	4,7%	3,2%	-3,0%	-2,2%	-1,1%	-0,6%	-0,5%	-0,5%	-0,5%	-0,5%	-0,6%	-0,6%	-0,9%	-0,8%	-0,5%	-0,8%	-0,8%	-0,2%	0,0%
$\Delta R3_{cp}$	13,8%	6,9%	4,6%	0,9%	0,0%	-0,6%	-1,0%	-0,9%	-0,8%	-0,7%	-0,6%	-0,3%	-0,8%	-0,2%	-0,2%	-0,6%	0,0%	0,0%	0,0%	0,0%
$\Delta R4_{cp}$	17,5%	8,8%	5,8%	0,6%	0,0%	-0,4%	0,0%	0,0%	0,0%	-0,2%	-0,2%	0,0%	-0,8%	-0,2%	-0,2%	-0,5%	0,0%	0,0%	0,0%	0,0%

Таблица В.7 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "6-34-Н-Н-Ми" к законопроекту "О полиции"

Квота для класса	Бюджет времени, Td/Td_{\max}																				
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1	
6-34-Н-Н-Ми	0	1	1	2	2	3	3	4	4	5	5	5	6	6	7	7	8	8	9	9	
Полнота отбора высказываний																					
R1	0	0,142857	0,142857	0,214286	0,214286	0,357143	0,357143	0,428571	0,428571	0,571429	0,571429	0,571429	0,642857	0,642857	0,714286	0,714286	0,857143	0,857143	1	1	
R2	0	0,152174	0,152174	0,195652	0,195652	0,23913	0,23913	0,282609	0,282609	0,434783	0,434783	0,434783	0,565217	0,565217	0,695652	0,695652	0,847826	0,847826	1	1	
R3	0	0,25	0,25	0,375	0,375	0,625	0,625	0,625	0,625	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,875	0,875	1	1	
R4	0	0,5	0,5	0,642857	0,642857	0,785714	0,785714	0,785714	0,785714	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,928571	0,928571	1	1	
R1 _{иcx}	0	0,142857	0,142857	0,285714	0,285714	0,357143	0,357143	0,5	0,5	0,642857	0,642857	0,642857	0,714286	0,714286	0,785714	0,785714	0,928571	0,928571	1	1	
R2 _{иcx}	0	0,152174	0,152174	0,304348	0,304348	0,347826	0,347826	0,5	0,5	0,652174	0,652174	0,652174	0,782609	0,782609	0,913043	0,913043	0,956522	0,956522	1	1	
R3 _{иcx}	0	0,25	0,25	0,375	0,375	0,5	0,5	0,625	0,625	0,75	0,75	0,75	0,75	0,75	0,75	0,75	1	1	1	1	
R4 _{иcx}	0	0,5	0,5	0,571429	0,571429	0,714286	0,714286	0,785714	0,785714	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	1	1	1	1	
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																					
$\Delta R1$	0,0%	14,3%	0,0%	7,1%	0,0%	14,3%	0,0%	7,1%	0,0%	14,3%	0,0%	0,0%	7,1%	0,0%	7,1%	0,0%	14,3%	0,0%	14,3%	0,0%	
$\Delta R2$	0,0%	15,2%	0,0%	4,3%	0,0%	4,3%	0,0%	4,3%	0,0%	15,2%	0,0%	0,0%	13,0%	0,0%	13,0%	0,0%	15,2%	0,0%	15,2%	0,0%	
$\Delta R3$	0,0%	25,0%	0,0%	12,5%	0,0%	25,0%	0,0%	0,0%	0,0%	12,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	12,5%	0,0%	12,5%	0,0%
$\Delta R4$	0,0%	50,0%	0,0%	14,3%	0,0%	14,3%	0,0%	0,0%	0,0%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	7,1%	0,0%	7,1%	0,0%
$\Delta R1_{\text{иcx}}$	0,0%	14,3%	0,0%	14,3%	0,0%	7,1%	0,0%	14,3%	0,0%	14,3%	0,0%	0,0%	7,1%	0,0%	7,1%	0,0%	14,3%	0,0%	7,1%	0,0%	
$\Delta R2_{\text{иcx}}$	0,0%	15,2%	0,0%	15,2%	0,0%	4,3%	0,0%	15,2%	0,0%	15,2%	0,0%	0,0%	13,0%	0,0%	13,0%	0,0%	4,3%	0,0%	4,3%	0,0%	
$\Delta R3_{\text{иcx}}$	0,0%	25,0%	0,0%	12,5%	0,0%	12,5%	0,0%	12,5%	0,0%	12,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	25,0%	0,0%	0,0%	0,0%
$\Delta R4_{\text{иcx}}$	0,0%	50,0%	0,0%	7,1%	0,0%	14,3%	0,0%	7,1%	0,0%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	14,3%	0,0%	0,0%	0,0%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																					
$(\Delta R1)_{\text{cp}}$	0,0%	7,1%	4,8%	5,4%	4,3%	6,0%	5,1%	5,4%	4,8%	5,7%	5,2%	4,8%	4,9%	4,6%	4,8%	4,5%	5,0%	4,8%	5,3%	5,0%	
$(\Delta R2)_{\text{cp}}$	0,0%	7,6%	5,1%	4,9%	3,9%	4,0%	3,4%	3,5%	3,1%	4,3%	4,0%	3,6%	4,3%	4,0%	4,6%	4,3%	5,0%	4,7%	5,3%	5,0%	
$(\Delta R3)_{\text{cp}}$	0,0%	12,5%	8,3%	9,4%	7,5%	10,4%	8,9%	7,8%	6,9%	7,5%	6,8%	6,3%	5,8%	5,4%	5,0%	4,7%	5,1%	4,9%	5,3%	5,0%	
$(\Delta R4)_{\text{cp}}$	0,0%	25,0%	16,7%	16,1%	12,9%	13,1%	11,2%	9,8%	8,7%	8,6%	7,8%	7,1%	6,6%	6,1%	5,7%	5,4%	5,5%	5,2%	5,3%	5,0%	
$(\Delta R1_{\text{иcx}})_{\text{cp}}$	0,0%	7,1%	4,8%	7,1%	5,7%	6,0%	5,1%	6,3%	5,6%	6,4%	5,8%	5,4%	5,5%	5,1%	5,2%	4,9%	5,5%	5,2%	5,3%	5,0%	
$(\Delta R2_{\text{иcx}})_{\text{cp}}$	0,0%	7,6%	5,1%	7,6%	6,1%	5,8%	5,0%	6,3%	5,6%	6,5%	5,9%	5,4%	6,0%	5,6%	6,1%	5,7%	5,6%	5,3%	5,3%	5,0%	
$(\Delta R3_{\text{иcx}})_{\text{cp}}$	0,0%	12,5%	8,3%	9,4%	7,5%	8,3%	7,1%	7,8%	6,9%	7,5%	6,8%	6,3%	5,8%	5,4%	5,0%	4,7%	5,9%	5,6%	5,3%	5,0%	
$(\Delta R4_{\text{иcx}})_{\text{cp}}$	0,0%	25,0%	16,7%	14,3%	11,4%	11,9%	10,2%	9,8%	8,7%	8,6%	7,8%	7,1%	6,6%	6,1%	5,7%	5,4%	5,9%	5,6%	5,3%	5,0%	
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																					
$\Delta R1_{\text{cp}}$	0,0%	0,0%	0,0%	-1,8%	-1,4%	0,0%	0,0%	-0,9%	-0,8%	-0,7%	-0,6%	-0,6%	-0,5%	-0,5%	-0,4%	-0,4%	-0,4%	0,0%	0,0%	0,0%	
$\Delta R2_{\text{cp}}$	0,0%	0,0%	0,0%	-2,7%	-2,2%	-1,8%	-1,6%	-2,7%	-2,4%	-2,2%	-2,0%	-1,8%	-1,7%	-1,6%	-1,4%	-1,4%	-0,6%	0,0%	0,0%	0,0%	
$\Delta R3_{\text{cp}}$	0,0%	0,0%	0,0%	0,0%	0,0%	2,1%	1,8%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,7%	-0,7%	0,0%	0,0%	
$\Delta R4_{\text{cp}}$	0,0%	0,0%	0,0%	1,8%	1,4%	1,2%	1,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,4%	-0,4%	0,0%	0,0%	

Таблица В.8 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "4-34-ГП-Н-Ми" к законопроекту "Об образовании"

Квота для класса	Бюджет времени, Td/Td_{\max}																					
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1		
4-34-ГП-Н-Ми	2	5	7	10	12	14	17	19	22	24	26	29	31	34	36	38	41	43	46	48		
Полнота отбора высказываний																						
R1	0,044776	0,149254	0,179104	0,238806	0,313433	0,343284	0,402985	0,432836	0,477612	0,522388	0,567164	0,61194	0,641791	0,686567	0,716418	0,776119	0,880597	0,925373	0,970149	1		
R2	0,057455	0,120727	0,147636	0,150545	0,179636	0,232	0,261091	0,313455	0,392	0,445091	0,502545	0,555636	0,582545	0,635636	0,688	0,750545	0,840727	0,869091	0,947636	1		
R3	0,086957	0,173913	0,217391	0,391304	0,565217	0,565217	0,695652	0,695652	0,695652	0,73913	0,73913	0,782609	0,826087	0,869565	0,869565	0,869565	0,956522	1	1	1		
R4	0,641791	0,701493	0,716418	0,776119	0,835821	0,835821	0,895522	0,895522	0,895522	0,910448	0,910448	0,925373	0,940299	0,955224	0,955224	0,955224	0,985075	1	1	1		
R1 _{исх}	0,029851	0,104478	0,134328	0,208955	0,253731	0,298507	0,373134	0,402985	0,507463	0,537313	0,567164	0,626866	0,656716	0,716418	0,746269	0,776119	0,835821	0,910448	0,970149	1		
R2 _{исх}	0,026909	0,085818	0,138182	0,192727	0,220364	0,248727	0,328727	0,355636	0,450182	0,502545	0,529455	0,613091	0,665455	0,723636	0,776	0,828364	0,882909	0,943273	0,998545	1		
R3 _{исх}	0,086957	0,217391	0,217391	0,347826	0,434783	0,521739	0,608696	0,652174	0,695652	0,695652	0,73913	0,73913	0,782609	0,782609	0,782609	0,826087	0,913043	0,913043	1			
R4 _{исх}	0,552239	0,686567	0,686567	0,731343	0,761194	0,80597	0,835821	0,850746	0,865672	0,865672	0,880597	0,880597	0,880597	0,895522	0,895522	0,910448	0,970149	0,970149	1			
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																						
$\Delta R1$	4,5%	10,4%	3,0%	6,0%	7,5%	3,0%	6,0%	3,0%	4,5%	4,5%	4,5%	4,5%	4,5%	3,0%	4,5%	3,0%	6,0%	10,4%	4,5%	4,5%	3,0%	
$\Delta R2$	5,7%	6,3%	2,7%	0,3%	2,9%	5,2%	2,9%	5,2%	7,9%	5,3%	5,7%	5,3%	2,7%	5,3%	5,2%	6,3%	9,0%	2,8%	7,9%	5,2%		
$\Delta R3$	8,7%	8,7%	4,3%	17,4%	17,4%	0,0%	13,0%	0,0%	0,0%	4,3%	0,0%	4,3%	4,3%	4,3%	4,3%	0,0%	0,0%	8,7%	4,3%	0,0%	0,0%	
$\Delta R4$	64,2%	6,0%	1,5%	6,0%	6,0%	0,0%	6,0%	0,0%	0,0%	1,5%	0,0%	1,5%	1,5%	1,5%	1,5%	0,0%	0,0%	3,0%	1,5%	0,0%	0,0%	
$\Delta R1_{\text{исх}}$	3,0%	7,5%	3,0%	7,5%	4,5%	4,5%	7,5%	3,0%	10,4%	3,0%	3,0%	6,0%	3,0%	6,0%	3,0%	3,0%	6,0%	7,5%	6,0%	3,0%	7,5%	
$\Delta R2_{\text{исх}}$	2,7%	5,9%	5,2%	5,5%	2,8%	2,8%	8,0%	2,7%	9,5%	5,2%	2,7%	8,4%	5,2%	5,2%	5,2%	5,5%	6,0%	5,5%	0,1%	6,0%	5,5%	
$\Delta R3_{\text{исх}}$	8,7%	13,0%	0,0%	13,0%	8,7%	8,7%	8,7%	4,3%	4,3%	0,0%	4,3%	0,0%	0,0%	4,3%	0,0%	4,3%	8,7%	0,0%	8,7%	0,0%	8,7%	
$\Delta R4_{\text{исх}}$	55,2%	13,4%	0,0%	4,5%	3,0%	4,5%	3,0%	1,5%	1,5%	0,0%	1,5%	0,0%	0,0%	1,5%	0,0%	1,5%	6,0%	0,0%	3,0%	6,0%	0,0%	
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																						
$(\Delta R1)_{\text{ср.}}$	4,5%	7,5%	6,0%	6,0%	6,3%	5,7%	5,8%	5,4%	5,3%	5,2%	5,2%	5,1%	4,9%	4,9%	4,8%	4,9%	5,2%	5,1%	5,1%	5,0%		
$(\Delta R2)_{\text{ср.}}$	5,7%	6,0%	4,9%	3,8%	3,6%	3,9%	3,7%	3,9%	4,4%	4,5%	4,6%	4,6%	4,5%	4,5%	4,6%	4,7%	4,9%	4,8%	5,0%	5,0%		
$(\Delta R3)_{\text{ср.}}$	8,7%	8,7%	7,2%	9,8%	11,3%	9,4%	9,9%	8,7%	7,7%	7,4%	6,7%	6,5%	6,4%	6,2%	5,8%	5,4%	5,6%	5,3%	5,0%			
$(\Delta R4)_{\text{ср.}}$	64,2%	35,1%	23,9%	19,4%	16,7%	13,9%	12,8%	11,2%	10,0%	9,1%	8,3%	7,7%	7,2%	6,8%	6,4%	6,0%	5,8%	5,6%	5,3%	5,0%		
$(\Delta R1_{\text{исх}})_{\text{ср.}}$	3,0%	5,2%	4,5%	5,2%	5,1%	5,0%	5,3%	5,0%	5,6%	5,4%	5,2%	5,2%	5,1%	5,1%	5,0%	4,9%	4,9%	5,1%	5,1%	5,0%		
$(\Delta R2_{\text{исх}})_{\text{ср.}}$	2,7%	4,3%	4,6%	4,8%	4,4%	4,1%	4,7%	4,4%	5,0%	5,0%	4,8%	5,1%	5,1%	5,2%	5,2%	5,2%	5,2%	5,3%	5,0%			
$(\Delta R3_{\text{исх}})_{\text{ср.}}$	8,7%	10,9%	7,2%	8,7%	8,7%	8,7%	8,7%	8,2%	7,7%	7,0%	6,7%	6,2%	5,7%	5,6%	5,2%	4,9%	4,9%	5,1%	4,8%	5,0%		
$(\Delta R4_{\text{исх}})_{\text{ср.}}$	55,2%	34,3%	22,9%	18,3%	15,2%	13,4%	11,9%	10,6%	9,6%	8,7%	8,0%	7,3%	6,8%	6,4%	6,0%	5,6%	5,4%	5,1%	5,0%			
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																						
$\Delta R1_{\text{ср.}}$	1,5%	2,2%	1,5%	0,7%	1,2%	0,7%	0,4%	0,4%	-0,3%	-0,1%	0,0%	-0,1%	-0,1%	-0,2%	-0,2%	0,0%	0,3%	0,1%	0,0%	0,0%		
$\Delta R2_{\text{ср.}}$	3,1%	1,7%	0,3%	-1,1%	-0,8%	-0,3%	-1,0%	-0,5%	-0,6%	-0,6%	-0,2%	-0,5%	-0,6%	-0,6%	-0,6%	-0,5%	-0,2%	-0,4%	-0,3%	0,0%		
$\Delta R3_{\text{ср.}}$	0,0%	-2,2%	0,0%	1,1%	2,6%	0,7%	1,2%	0,5%	0,0%	0,4%	0,0%	0,4%	0,7%	0,6%	0,6%	0,5%	0,8%	0,5%	0,5%	0,0%		
$\Delta R4_{\text{ср.}}$	9,0%	0,7%	1,0%	1,1%	1,5%	0,5%	0,9%	0,6%	0,3%	0,4%	0,3%	0,4%	0,5%	0,4%	0,4%	0,4%	0,4%	0,2%	0,2%	0,0%		

Таблица В.9 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

Квота для класса	Бюджет времени, Td/Td_{\max}																			
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
1-11-ОД-Н-Ми	1	2	3	4	5	6	7	8	9	10	10	11	12	13	14	15	16	17	18	19
Полнота отбора высказываний																				
R1	0,162162	0,189189	0,243243	0,27027	0,297297	0,351351	0,378378	0,432432	0,513514	0,567568	0,567568	0,594595	0,702703	0,72973	0,783784	0,810811	0,837838	0,945946	0,972973	1
R2	0,134021	0,144433	0,164948	0,185567	0,195876	0,216495	0,226804	0,309278	0,43299	0,546392	0,546392	0,556701	0,680412	0,731959	0,752577	0,804124	0,824742	0,927835	0,948454	1
R3	0,26087	0,304348	0,391304	0,434783	0,478261	0,565217	0,608696	0,608696	0,695652	0,695652	0,73913	0,826087	0,826087	0,913043	0,913043	0,913043	1	1	1	1
R4	0,351351	0,378378	0,432432	0,486486	0,513514	0,567568	0,594595	0,594595	0,756757	0,756757	0,756757	0,783784	0,891892	0,891892	0,945946	0,945946	0,945946	1	1	1
R1 _{исх}	0,054054	0,081081	0,135135	0,162162	0,189189	0,216216	0,324324	0,378378	0,540541	0,648649	0,648649	0,702703	0,72973	0,756757	0,783784	0,810811	0,864865	0,891892	0,972973	1
R2 _{исх}	0,020619	0,041237	0,061856	0,113402	0,123711	0,14433	0,247423	0,268041	0,402062	0,525773	0,525773	0,608247	0,659794	0,71134	0,731959	0,742268	0,85567	0,865979	0,989691	1
R3 _{исх}	0,086957	0,130435	0,217391	0,26087	0,304348	0,347826	0,521739	0,608696	0,826087	0,826087	0,826087	0,826087	0,826087	0,826087	0,869565	0,869565	0,913043	0,956522	1	1
R4 _{исх}	0,054054	0,108108	0,162162	0,297297	0,324324	0,378378	0,648649	0,702703	0,891892	0,891892	0,891892	0,891892	0,891892	0,918919	0,918919	0,945946	0,972973	1	1	1
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1$	16,2%	2,7%	5,4%	2,7%	2,7%	5,4%	2,7%	5,4%	8,1%	5,4%	0,0%	2,7%	10,8%	2,7%	5,4%	2,7%	2,7%	10,8%	2,7%	2,7%
$\Delta R2$	13,4%	1,0%	2,1%	2,1%	1,0%	2,1%	1,0%	8,2%	12,4%	11,3%	0,0%	1,0%	12,4%	5,2%	2,1%	5,2%	2,1%	10,3%	2,1%	5,2%
$\Delta R3$	26,1%	4,3%	8,7%	4,3%	4,3%	8,7%	4,3%	0,0%	8,7%	0,0%	0,0%	4,3%	8,7%	0,0%	8,7%	0,0%	0,0%	8,7%	0,0%	0,0%
$\Delta R4$	35,1%	2,7%	5,4%	5,4%	2,7%	5,4%	2,7%	0,0%	16,2%	0,0%	0,0%	2,7%	10,8%	0,0%	5,4%	0,0%	0,0%	5,4%	0,0%	0,0%
$\Delta R1_{\text{исх}}$	5,4%	2,7%	5,4%	2,7%	2,7%	2,7%	10,8%	5,4%	16,2%	10,8%	0,0%	5,4%	2,7%	2,7%	2,7%	5,4%	2,7%	8,1%	2,7%	2,7%
$\Delta R2_{\text{исх}}$	2,1%	2,1%	2,1%	5,2%	1,0%	2,1%	10,3%	2,1%	13,4%	12,4%	0,0%	8,2%	5,2%	5,2%	2,1%	1,0%	11,3%	1,0%	12,4%	1,0%
$\Delta R3_{\text{исх}}$	8,7%	4,3%	8,7%	4,3%	4,3%	4,3%	17,4%	8,7%	21,7%	0,0%	0,0%	0,0%	0,0%	0,0%	4,3%	0,0%	4,3%	4,3%	4,3%	4,3%
$\Delta R4_{\text{исх}}$	5,4%	5,4%	5,4%	13,5%	2,7%	5,4%	27,0%	5,4%	18,9%	0,0%	0,0%	0,0%	0,0%	0,0%	2,7%	0,0%	2,7%	2,7%	2,7%	2,7%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$(\Delta R1)_{\text{ср.}}$	16,2%	9,5%	8,1%	6,8%	5,9%	5,9%	5,4%	5,4%	5,7%	5,7%	5,2%	5,0%	5,4%	5,2%	5,1%	4,9%	5,3%	5,1%	5,0%	5,0%
$(\Delta R2)_{\text{ср.}}$	13,4%	7,2%	5,5%	4,6%	3,9%	3,6%	3,2%	3,9%	4,8%	5,5%	5,0%	4,6%	5,2%	5,2%	5,0%	4,9%	5,2%	5,0%	5,0%	5,0%
$(\Delta R3)_{\text{ср.}}$	26,1%	15,2%	13,0%	10,9%	9,6%	9,4%	8,7%	7,6%	7,7%	7,0%	6,3%	6,2%	6,4%	5,9%	6,1%	5,7%	5,4%	5,6%	5,3%	5,0%
$(\Delta R4)_{\text{ср.}}$	35,1%	18,9%	14,4%	12,2%	10,3%	9,5%	8,5%	7,4%	8,4%	7,6%	6,9%	6,5%	6,9%	6,4%	6,3%	5,9%	5,6%	5,3%	5,0%	5,0%
$(\Delta R1_{\text{исх}})_{\text{ср.}}$	5,4%	4,1%	4,5%	4,1%	3,8%	3,6%	4,6%	4,7%	6,0%	6,5%	5,9%	5,9%	5,6%	5,4%	5,2%	5,1%	5,0%	5,1%	5,0%	5,0%
$(\Delta R2_{\text{исх}})_{\text{ср.}}$	2,1%	2,1%	2,1%	2,8%	2,5%	2,4%	3,5%	3,4%	4,5%	5,3%	4,8%	5,1%	5,1%	4,9%	4,6%	5,0%	4,8%	5,2%	5,0%	5,0%
$(\Delta R3_{\text{исх}})_{\text{ср.}}$	8,7%	6,5%	7,2%	6,5%	6,1%	5,8%	7,5%	7,6%	9,2%	8,3%	7,5%	6,9%	6,4%	5,9%	5,5%	5,4%	5,1%	5,0%	5,0%	5,0%
$(\Delta R4_{\text{исх}})_{\text{ср.}}$	5,4%	5,4%	5,4%	7,4%	6,5%	6,3%	9,3%	8,8%	9,9%	8,9%	8,1%	7,4%	6,9%	6,4%	5,9%	5,7%	5,4%	5,3%	5,1%	5,0%
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1_{\text{ср.}}$	10,8%	5,4%	3,6%	2,7%	2,2%	2,3%	0,8%	0,7%	-0,3%	-0,8%	-0,7%	-0,9%	-0,2%	-0,2%	0,0%	0,0%	-0,2%	0,3%	0,0%	0,0%
$\Delta R2_{\text{ср.}}$	11,3%	5,2%	3,4%	1,8%	1,4%	1,2%	-0,3%	0,5%	0,3%	0,2%	0,2%	-0,4%	0,2%	0,1%	0,1%	0,4%	-0,2%	0,3%	-0,2%	0,0%
$\Delta R3_{\text{ср.}}$	17,4%	8,7%	5,8%	4,3%	3,5%	3,6%	1,2%	0,0%	-1,4%	-1,3%	-1,2%	-0,7%	0,0%	0,0%	0,6%	0,3%	0,3%	0,5%	0,2%	0,0%
$\Delta R4_{\text{ср.}}$	29,7%	13,5%	9,0%	4,7%	3,8%	3,2%	-0,8%	-1,4%	-1,5%	-1,4%	-1,2%	-0,9%	0,0%	0,0%	0,4%	0,2%	0,2%	0,3%	0,1%	0,0%

Таблица В.10 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "1-1-ОД-Н-Ми" к законопроекту "Об образовании"

Квота для класса	Бюджет времени, Td/Td_{\max}																			
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
1-1-ОД-Н-Ми	1	1	2	2	3	4	4	5	5	6	7	7	8	8	9	10	10	11	11	12
Полнота отбора высказываний																				
R1	0,125	0,125	0,25	0,25	0,3125	0,375	0,375	0,4375	0,4375	0,5	0,5625	0,5625	0,625	0,625	0,6875	0,75	0,75	0,875	0,875	1
R2	0,111111	0,111111	0,277778	0,277778	0,333333	0,388889	0,388889	0,444444	0,444444	0,5	0,555556	0,555556	0,611111	0,611111	0,666667	0,722222	0,722222	0,888889	0,888889	1
R3	0,133333	0,133333	0,266667	0,266667	0,333333	0,4	0,4	0,466667	0,466667	0,533333	0,6	0,6	0,666667	0,666667	0,733333	0,8	0,8	0,866667	0,866667	1
R4	0,125	0,125	0,3125	0,3125	0,375	0,4375	0,4375	0,5	0,5	0,5625	0,625	0,625	0,6875	0,6875	0,75	0,8125	0,8125	0,875	0,875	1
R1 _{исх}	0,125	0,125	0,1875	0,1875	0,25	0,3125	0,3125	0,4375	0,4375	0,5	0,5625	0,5625	0,6875	0,6875	0,75	0,875	0,875	0,9375	0,9375	1
R2 _{исх}	0,111111	0,111111	0,166667	0,166667	0,222222	0,277778	0,277778	0,444444	0,444444	0,5	0,555556	0,555556	0,666667	0,666667	0,722222	0,888889	0,888889	0,944444	0,944444	1
R3 _{исх}	0,133333	0,133333	0,2	0,2	0,266667	0,333333	0,333333	0,466667	0,466667	0,533333	0,6	0,6	0,733333	0,733333	0,8	0,866667	0,866667	0,933333	0,933333	1
R4 _{исх}	0,125	0,125	0,1875	0,1875	0,25	0,3125	0,3125	0,5	0,5	0,5625	0,625	0,625	0,75	0,75	0,8125	0,875	0,875	0,9375	0,9375	1
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1$	12,5%	0,0%	12,5%	0,0%	6,3%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	12,5%	0,0%	12,5%	0,0%	12,5%
$\Delta R2$	11,1%	0,0%	16,7%	0,0%	5,6%	5,6%	0,0%	5,6%	0,0%	5,6%	0,0%	5,6%	0,0%	5,6%	0,0%	16,7%	0,0%	11,1%	0,0%	11,1%
$\Delta R3$	13,3%	0,0%	13,3%	0,0%	6,7%	6,7%	0,0%	6,7%	0,0%	6,7%	0,0%	6,7%	0,0%	6,7%	0,0%	6,7%	0,0%	6,7%	0,0%	13,3%
$\Delta R4$	12,5%	0,0%	18,8%	0,0%	6,3%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	6,3%	0,0%	12,5%
$\Delta R1_{исх}$	12,5%	0,0%	6,3%	0,0%	6,3%	6,3%	0,0%	12,5%	0,0%	6,3%	6,3%	0,0%	12,5%	0,0%	6,3%	12,5%	0,0%	6,3%	0,0%	6,3%
$\Delta R2_{исх}$	11,1%	0,0%	5,6%	0,0%	5,6%	5,6%	0,0%	16,7%	0,0%	5,6%	5,6%	0,0%	11,1%	0,0%	5,6%	16,7%	0,0%	5,6%	0,0%	5,6%
$\Delta R3_{исх}$	13,3%	0,0%	6,7%	0,0%	6,7%	6,7%	0,0%	13,3%	0,0%	6,7%	6,7%	0,0%	13,3%	0,0%	6,7%	6,7%	0,0%	6,7%	0,0%	6,7%
$\Delta R4_{исх}$	12,5%	0,0%	6,3%	0,0%	6,3%	6,3%	0,0%	18,8%	0,0%	6,3%	6,3%	0,0%	12,5%	0,0%	6,3%	6,3%	0,0%	6,3%	0,0%	6,3%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$(\Delta R1)_{cp}$	12,5%	6,3%	8,3%	6,3%	6,3%	6,3%	5,4%	5,5%	4,9%	5,0%	5,1%	4,7%	4,8%	4,5%	4,6%	4,7%	4,4%	4,9%	4,6%	5,0%
$(\Delta R2)_{cp}$	11,1%	5,6%	9,3%	6,9%	6,7%	6,5%	5,6%	5,6%	4,9%	5,0%	5,1%	4,6%	4,7%	4,4%	4,4%	4,5%	4,2%	4,9%	4,7%	5,0%
$(\Delta R3)_{cp}$	13,3%	6,7%	8,9%	6,7%	6,7%	6,7%	5,7%	5,8%	5,2%	5,3%	5,5%	5,0%	5,1%	4,8%	4,9%	5,0%	4,7%	4,8%	4,6%	5,0%
$(\Delta R4)_{cp}$	12,5%	6,3%	10,4%	7,8%	7,5%	7,3%	6,3%	6,3%	5,6%	5,6%	5,7%	5,2%	5,3%	4,9%	5,0%	5,1%	4,8%	4,9%	4,6%	5,0%
$(\Delta R1_{исх})_{cp}$	12,5%	6,3%	6,3%	4,7%	5,0%	5,2%	4,5%	5,5%	4,9%	5,0%	5,1%	4,7%	5,3%	4,9%	5,0%	5,5%	5,1%	5,2%	4,9%	5,0%
$(\Delta R2_{исх})_{cp}$	11,1%	5,6%	5,6%	4,2%	4,4%	4,6%	4,0%	5,6%	4,9%	5,0%	5,1%	4,6%	5,1%	4,8%	4,8%	5,6%	5,2%	5,2%	5,0%	5,0%
$(\Delta R3_{исх})_{cp}$	13,3%	6,7%	6,7%	5,0%	5,3%	5,6%	4,8%	5,8%	5,2%	5,3%	5,5%	5,0%	5,6%	5,2%	5,3%	5,4%	5,1%	5,2%	4,9%	5,0%
$(\Delta R4_{исх})_{cp}$	12,5%	6,3%	6,3%	4,7%	5,0%	5,2%	4,5%	6,3%	5,6%	5,6%	5,7%	5,2%	5,8%	5,4%	5,5%	5,1%	5,2%	4,9%	5,0%	5,0%
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1_{cp}$	0,0%	0,0%	2,1%	1,6%	1,3%	1,0%	0,9%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,5%	-0,4%	-0,4%	-0,8%	-0,7%	-0,3%	-0,3%	0,0%
$\Delta R2_{cp}$	0,0%	0,0%	3,7%	2,8%	2,2%	1,9%	1,6%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,4%	-0,4%	-0,4%	-0,4%	-1,0%	-0,3%	-0,3%	0,0%
$\Delta R3_{cp}$	0,0%	0,0%	2,2%	1,7%	1,3%	1,1%	1,0%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,5%	-0,5%	-0,4%	-0,4%	-0,4%	-0,4%	-0,4%	0,0%
$\Delta R4_{cp}$	0,0%	0,0%	4,2%	3,1%	2,5%	2,1%	1,8%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,5%	-0,4%	-0,4%	-0,4%	-0,4%	-0,3%	-0,3%	0,0%

ПРИЛОЖЕНИЕ Г
РАСЧЕТ СРЕДНЕГО ТЕМПА ПРИРОСТА ПОЛНОТЫ ОТБОРА ВЫСКАЗЫВАНИЙ
В ВЫБОРКЕ ПОЛЕЗНЫХ КОММЕНТАРИЕВ

Таблица Г.1 – Расчет пошагового накопления полноты отбора высказываний для отранжированного по полезности множества комментариев класса "7-42-Н-В-Ми" к законопроекту "О полиции"

u_{v,j_v}	Ид. комм.	m_v																													Кол. выск.	Накоплено высказываний				Полнота					
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29		б/вес, б/повт	вес, б/повт	б/вес, повт	вес, повт	R1	R2	R3	R4		
0,025762	2839	1	1			1	1	1	1	1																					7	7	15	7	15	0,175	0,202703	0,241379	0,375		
0,016452	13331																																1	8	16	8	16	0,2	0,216216	0,275862	0,4
0,010941	16593																																1	9	19	9	19	0,225	0,256757	0,310345	0,475
0,00931	2715							1																									1	10	20	10	20	0,25	0,27027	0,344828	0,5
0,007142	5927										1																						1	11	21	11	21	0,275	0,283784	0,37931	0,525
0,007142	11828																		1														1	12	22	12	22	0,3	0,297297	0,413793	0,55
0,005152	15105	1																															1	13	26	12	22	0,325	0,351351	0,413793	0,55
0,004158	359	1	1	1																												3	16	34	13	23	0,4	0,459459	0,448276	0,575	
0,004158	10327																		1													1	17	35	14	24	0,425	0,472973	0,482759	0,6	
0,004158	12232																		1													1	18	37	15	26	0,45	0,5	0,517241	0,65	
0,004158	15430																		1													1	19	39	15	26	0,475	0,527027	0,517241	0,65	
0,004158	16544	1																													1	20	43	15	26	0,5	0,581081	0,517241	0,65		
0,002984	1779			1																											1	21	44	16	27	0,525	0,594595	0,551724	0,675		
0,000995	5639																		1													1	22	45	17	28	0,55	0,608108	0,586207	0,7	
0,000995	10872																														1	23	46	18	29	0,575	0,621622	0,62069	0,725		
0,000995	14937																														2	25	50	19	30	0,625	0,675676	0,655172	0,75		
0,000995	16281																														2	27	52	21	32	0,675	0,702703	0,724138	0,8		
0,000995	19074																														1	1	28	53	22	0,7	0,716216	0,758621	0,825		
0	1658			1																											1	29	54	23	34	0,725	0,72973	0,793103	0,85		
0	3875																		1												1	30	57	23	34	0,75	0,77027	0,793103	0,85		
0	7041																		1												2	32	61	24	35	0,8	0,824324	0,827586	0,875		
0	7109																		1												1	33	62	25	36	0,825	0,837838	0,862069	0,9		
0	13825		1																												2	35	67	25	36	0,875	0,905405	0,862069	0,9		
0	13869																														1	36	68	26	37	0,9	0,918919	0,896552	0,925		
0	15920																														1	37	69	27	38	0,925	0,932432	0,931034	0,95		
0	16327																														1	38	72	27	38	0,95	0,972973	0,931034	0,95		
0	16736																														1	39	73	28	39	0,975	0,986486	0,965517	0,975		
0	18525																														1	40	74	29	40	1	1	1	1		
Вес высказыв.		1	4	3	1	1	1	3	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	40									

Таблица Г.2 – Расчет пошагового накопления полноты отбора высказываний для отранжированного по полезности множества комментариев класса "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

u_v, j_v	Ид. комм.	m_v																							Кол. высск.	Накоплено высказываний				Полнота				
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23		б/вес, б/повт	вес, б/повт	б/вес, повт	вес, повт	R1	R2	R3	R4	
0,062984	7754						1			1														1	3	3	12	3	12	0,081081	0,123711	0,130435	0,324324	
0,027587	2823									1															6	9	25	8	19	0,243243	0,257732	0,347826	0,513514	
0,026519	2956									1	1	1	1	1	1									4	13	37	10	23	0,351351	0,381443	0,434783	0,621622		
0,02545	510					1	1																		2	15	39	12	25	0,405405	0,402062	0,521739	0,675676	
0,019104	1052								1																1	16	40	13	26	0,432432	0,412371	0,565217	0,702703	
0,016967	190	1	1																						2	18	42	15	28	0,486486	0,43299	0,652174	0,756757	
0,014961	7077																								1	19	43	16	29	0,513514	0,443299	0,695652	0,783784	
0,009618	8095																								1	1	20	44	17	30	0,540541	0,453608	0,73913	0,810811
0,009018	1872							1																	1	21	46	17	30	0,567568	0,474227	0,73913	0,810811	
0,009018	3615								1																2	23	54	17	30	0,621622	0,556701	0,73913	0,810811	
0,009018	3616									1															1	24	59	17	30	0,648649	0,608247	0,73913	0,810811	
0,008483	3925									1															1	25	64	17	30	0,675676	0,659794	0,73913	0,810811	
0,008483	4129																								1	26	65	18	31	0,702703	0,670103	0,782609	0,837838	
0,002672	5177									1			1												2	28	76	18	31	0,756757	0,783505	0,782609	0,837838	
0,001069	2343										1	1	1	1										4	32	86	20	33	0,864865	0,886598	0,869565	0,891892		
0,000534	513										1														1	33	91	20	33	0,891892	0,938144	0,869565	0,891892	
0,000534	4006										1														1	34	93	21	35	0,918919	0,958763	0,913043	0,945946	
0	468																								1	35	95	21	35	0,945946	0,979381	0,913043	0,945946	
0	2346																								2	37	97	23	37	1	1	1	1	
Вес высказыв.		1	1	2	1	1	5	1	2	1	6	1	2	1	1	2	1	1	2	1	1	1	1	37										

Таблица Г.3 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "3-12-Од-В-Ми" к законопроекту "О полиции"

Квота для класса	Бюджет времени, Td/Td_{\max}																			
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
3-12-Од-Ми	1	2	6	6	10	16	31	39	54	101	86	102	117	133	150	157	193	200	203	203
3-12-Од-В-Ми	0	0	2	2	3	5	10	12	17	25	21	26	30	35	42	43	54	56	58	58
Полнота отбора высказываний																				
R1	0	0	0,028571	0,028571	0,057143	0,142857	0,214286	0,3	0,371429	0,5	0,442857	0,514286	0,571429	0,642857	0,742857	0,757143	0,942857	0,971429	1	1
R2	0	0	0,015873	0,015873	0,055556	0,119048	0,18254	0,277778	0,333333	0,444444	0,380952	0,452381	0,515873	0,587302	0,690476	0,698413	0,936508	0,960317	1	1
R3	0	0	0,04	0,04	0,08	0,2	0,28	0,38	0,48	0,6	0,56	0,62	0,68	0,74	0,82	0,84	0,98	1	1	1
R4	0	0	0,028571	0,028571	0,1	0,214286	0,285714	0,428571	0,528571	0,642857	0,585714	0,657143	0,728571	0,785714	0,842857	0,857143	0,985714	1	1	1
R1 _{исх}	0	0	0,028571	0,028571	0,042857	0,071429	0,142857	0,171429	0,314286	0,428571	0,371429	0,442857	0,514286	0,657143	0,757143	0,771429	0,928571	0,971429	1	1
R2 _{исх}	0	0	0,031746	0,031746	0,039683	0,063492	0,111111	0,134921	0,301587	0,396825	0,34127	0,404762	0,484127	0,650794	0,730159	0,746032	0,944444	0,984127	1	1
R3 _{исх}	0	0	0,04	0,04	0,06	0,1	0,2	0,24	0,4	0,52	0,48	0,54	0,6	0,7	0,82	0,82	0,92	0,96	1	1
R4 _{исх}	0	0	0,057143	0,057143	0,071429	0,114286	0,2	0,242857	0,442857	0,557143	0,514286	0,571429	0,657143	0,742857	0,857143	0,857143	0,942857	0,971429	1	1
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1$	0,0%	0,0%	2,9%	0,0%	2,9%	8,6%	7,1%	8,6%	7,1%	12,9%	-5,7%	7,1%	5,7%	7,1%	10,0%	1,4%	18,6%	2,9%	2,9%	0,0%
$\Delta R2$	0,0%	0,0%	1,6%	0,0%	4,0%	6,3%	6,3%	9,5%	5,6%	11,1%	-6,3%	7,1%	6,3%	7,1%	10,3%	0,8%	23,8%	2,4%	4,0%	0,0%
$\Delta R3$	0,0%	0,0%	4,0%	0,0%	4,0%	12,0%	8,0%	10,0%	10,0%	12,0%	-4,0%	6,0%	6,0%	6,0%	8,0%	2,0%	14,0%	2,0%	0,0%	0,0%
$\Delta R4$	0,0%	0,0%	2,9%	0,0%	7,1%	11,4%	7,1%	14,3%	10,0%	11,4%	-5,7%	7,1%	7,1%	5,7%	5,7%	1,4%	12,9%	1,4%	0,0%	0,0%
$\Delta R1_{исх}$	0,0%	0,0%	2,9%	0,0%	1,4%	2,9%	7,1%	2,9%	14,3%	11,4%	-5,7%	7,1%	7,1%	14,3%	10,0%	1,4%	15,7%	4,3%	2,9%	0,0%
$\Delta R2_{исх}$	0,0%	0,0%	3,2%	0,0%	0,8%	2,4%	4,8%	2,4%	16,7%	9,5%	-5,6%	6,3%	7,9%	16,7%	7,9%	1,6%	19,8%	4,0%	1,6%	0,0%
$\Delta R3_{исх}$	0,0%	0,0%	4,0%	0,0%	2,0%	4,0%	10,0%	4,0%	16,0%	12,0%	-4,0%	6,0%	6,0%	10,0%	12,0%	0,0%	10,0%	4,0%	4,0%	0,0%
$\Delta R4_{исх}$	0,0%	0,0%	5,7%	0,0%	1,4%	4,3%	8,6%	4,3%	20,0%	11,4%	-4,3%	5,7%	8,6%	8,6%	11,4%	0,0%	8,6%	2,9%	2,9%	0,0%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$(\Delta R1)_{cp}$	0,0%	0,0%	1,0%	0,7%	1,1%	2,4%	3,1%	3,8%	4,1%	5,0%	4,0%	4,3%	4,4%	4,6%	5,0%	4,7%	5,5%	5,4%	5,3%	5,0%
$(\Delta R2)_{cp}$	0,0%	0,0%	0,5%	0,4%	1,1%	2,0%	2,6%	3,5%	3,7%	4,4%	3,5%	3,8%	4,0%	4,2%	4,6%	4,4%	5,5%	5,3%	5,3%	5,0%
$(\Delta R3)_{cp}$	0,0%	0,0%	1,3%	1,0%	1,6%	3,3%	4,0%	4,8%	5,3%	6,0%	5,1%	5,2%	5,2%	5,3%	5,5%	5,3%	5,8%	5,6%	5,3%	5,0%
$(\Delta R4)_{cp}$	0,0%	0,0%	1,0%	0,7%	2,0%	3,6%	4,1%	5,4%	5,9%	6,4%	5,3%	5,5%	5,6%	5,6%	5,4%	5,8%	5,6%	5,3%	5,0%	5,0%
$(\Delta R1_{исх})_{cp}$	0,0%	0,0%	1,0%	0,7%	0,9%	1,2%	2,0%	2,1%	3,5%	4,3%	3,4%	3,7%	4,0%	4,7%	5,0%	4,8%	5,5%	5,4%	5,3%	5,0%
$(\Delta R2_{исх})_{cp}$	0,0%	0,0%	1,1%	0,8%	0,8%	1,1%	1,6%	1,7%	3,4%	4,0%	3,1%	3,4%	3,7%	4,6%	4,9%	4,7%	5,6%	5,5%	5,3%	5,0%
$(\Delta R3_{исх})_{cp}$	0,0%	0,0%	1,3%	1,0%	1,2%	1,7%	2,9%	3,0%	4,4%	5,2%	4,4%	4,5%	4,6%	5,0%	5,5%	5,1%	5,4%	5,3%	5,3%	5,0%
$(\Delta R4_{исх})_{cp}$	0,0%	0,0%	1,9%	1,4%	1,4%	1,9%	2,9%	3,0%	4,9%	5,6%	4,7%	4,8%	5,1%	5,3%	5,7%	5,4%	5,5%	5,4%	5,3%	5,0%
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1_{cp}$	0,0%	0,0%	0,0%	0,0%	0,3%	1,2%	1,0%	1,6%	0,6%	0,7%	0,6%	0,6%	0,4%	-0,1%	-0,1%	-0,1%	0,1%	0,1%	0,0%	0,0%
$\Delta R2_{cp}$	0,0%	0,0%	-0,5%	-0,4%	0,3%	0,9%	1,0%	1,8%	0,4%	0,5%	0,4%	0,4%	0,2%	-0,5%	-0,3%	-0,3%	0,0%	-0,1%	0,0%	0,0%
$\Delta R3_{cp}$	0,0%	0,0%	0,0%	0,0%	0,4%	1,7%	1,1%	1,8%	0,9%	0,8%	0,7%	0,7%	0,6%	0,3%	0,0%	0,1%	0,4%	0,2%	0,0%	0,0%
$\Delta R4_{cp}$	0,0%	0,0%	-1,0%	-0,7%	0,6%	1,7%	1,2%	2,3%	1,0%	0,9%	0,6%	0,7%	0,5%	0,3%	-0,1%	0,0%	0,3%	0,2%	0,0%	0,0%

Таблица Г.4 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "7-42-Н-В-Ми" к законопроекту "О полиции"

Таблица Г.5 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "6-34-Н-Н-Ми" к законопроекту "О полиции"

Квота для класса	Бюджет времени, Td/Td_{\max}																			
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
6-34-Н-Ми	1	2	2	3	6	7	7	9	10	9	10	12	13	14	12	12	13	18	18	18
6-34-Н-Н-Ми	0	1	1	1	2	3	3	5	6	5	6	7	7	7	7	7	9	9	9	9
Полнота отбора высказываний																				
R1	0	0,142857	0,142857	0,142857	0,285714	0,428571	0,428571	0,642857	0,785714	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	
R2	0	0,152174	0,152174	0,152174	0,304348	0,347826	0,347826	0,630435	0,782609	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	
R3	0	0,25	0,25	0,25	0,375	0,625	0,625	0,75	0,875	0,75	0,875	0,875	0,875	0,875	0,875	0,875	0,875	0,875	0,875	0,875
R4	0	0,5	0,5	0,5	0,571429	0,714286	0,714286	0,785714	0,857143	0,785714	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	
R1 _{нсx}	0	0,142857	0,142857	0,142857	0,285714	0,357143	0,357143	0,642857	0,714286	0,642857	0,714286	0,785714	0,785714	0,785714	0,785714	0,785714	0,785714	0,785714	0,785714	0,785714
R2 _{нсx}	0	0,152174	0,152174	0,152174	0,304348	0,347826	0,347826	0,652174	0,782609	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043	0,913043
R3 _{нсx}	0	0,25	0,25	0,25	0,375	0,5	0,5	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75	0,75
R4 _{нсx}	0	0,5	0,5	0,5	0,571429	0,714286	0,714286	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143	0,857143
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1$	0,0%	14,3%	0,0%	0,0%	14,3%	14,3%	0,0%	21,4%	14,3%	-14,3%	14,3%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%	14,3%	0,0%	0,0%
$\Delta R2$	0,0%	15,2%	0,0%	0,0%	15,2%	4,3%	0,0%	28,3%	15,2%	-15,2%	15,2%	13,0%	0,0%	0,0%	0,0%	0,0%	0,0%	8,7%	0,0%	0,0%
$\Delta R3$	0,0%	25,0%	0,0%	0,0%	12,5%	25,0%	0,0%	12,5%	12,5%	-12,5%	12,5%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	12,5%	0,0%	0,0%
$\Delta R4$	0,0%	50,0%	0,0%	0,0%	7,1%	14,3%	0,0%	7,1%	7,1%	-7,1%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	14,3%	0,0%	0,0%
$\Delta R1_{\text{нсx}}$	0,0%	14,3%	0,0%	0,0%	14,3%	7,1%	0,0%	28,6%	7,1%	-7,1%	7,1%	7,1%	0,0%	0,0%	0,0%	0,0%	0,0%	21,4%	0,0%	0,0%
$\Delta R2_{\text{нсx}}$	0,0%	15,2%	0,0%	0,0%	15,2%	4,3%	0,0%	30,4%	13,0%	-13,0%	13,0%	13,0%	0,0%	0,0%	0,0%	0,0%	0,0%	8,7%	0,0%	0,0%
$\Delta R3_{\text{нсx}}$	0,0%	25,0%	0,0%	0,0%	12,5%	12,5%	0,0%	25,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	25,0%	0,0%	0,0%
$\Delta R4_{\text{нсx}}$	0,0%	50,0%	0,0%	0,0%	7,1%	14,3%	0,0%	14,3%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	14,3%	0,0%	0,0%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$(\Delta R1)_{\text{cp.}}$	0,0%	7,1%	4,8%	3,6%	5,7%	7,1%	6,1%	8,0%	8,7%	6,4%	7,1%	7,1%	6,6%	6,1%	5,7%	5,4%	5,0%	5,6%	5,3%	5,0%
$(\Delta R2)_{\text{cp.}}$	0,0%	7,6%	5,1%	3,8%	6,1%	5,8%	5,0%	7,9%	8,7%	6,3%	7,1%	7,6%	7,0%	6,5%	6,1%	5,7%	5,4%	5,6%	5,3%	5,0%
$(\Delta R3)_{\text{cp.}}$	0,0%	12,5%	8,3%	6,3%	7,5%	10,4%	8,9%	9,4%	9,7%	7,5%	8,0%	7,3%	6,7%	6,3%	5,8%	5,5%	5,1%	5,6%	5,3%	5,0%
$(\Delta R4)_{\text{cp.}}$	0,0%	25,0%	16,7%	12,5%	11,4%	11,9%	10,2%	9,8%	9,5%	7,9%	7,8%	7,1%	6,6%	6,1%	5,7%	5,4%	5,0%	5,6%	5,3%	5,0%
$(\Delta R1_{\text{нсx}})_{\text{cp.}}$	0,0%	7,1%	4,8%	3,6%	5,7%	6,0%	5,1%	8,0%	7,9%	6,4%	6,5%	6,5%	6,0%	5,6%	5,2%	4,9%	4,6%	5,6%	5,3%	5,0%
$(\Delta R2_{\text{нсx}})_{\text{cp.}}$	0,0%	7,6%	5,1%	3,8%	6,1%	5,8%	5,0%	8,2%	8,7%	6,5%	7,1%	7,6%	7,0%	6,5%	6,1%	5,7%	5,4%	5,6%	5,3%	5,0%
$(\Delta R3_{\text{нсx}})_{\text{cp.}}$	0,0%	12,5%	8,3%	6,3%	7,5%	8,3%	7,1%	9,4%	8,3%	7,5%	6,8%	6,3%	5,8%	5,4%	5,0%	4,7%	4,4%	5,6%	5,3%	5,0%
$(\Delta R4_{\text{нсx}})_{\text{cp.}}$	0,0%	25,0%	16,7%	12,5%	11,4%	11,9%	10,2%	10,7%	9,5%	8,6%	7,8%	7,1%	6,6%	6,1%	5,7%	5,4%	5,0%	5,6%	5,3%	5,0%
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1_{\text{cp.}}$	0,0%	0,0%	0,0%	0,0%	0,0%	1,2%	1,0%	0,0%	0,8%	0,0%	0,6%	0,6%	0,5%	0,5%	0,5%	0,4%	0,4%	0,0%	0,0%	0,0%
$\Delta R2_{\text{cp.}}$	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,3%	0,0%	-0,2%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
$\Delta R3_{\text{cp.}}$	0,0%	0,0%	0,0%	0,0%	0,0%	2,1%	1,8%	0,0%	1,4%	0,0%	1,1%	1,0%	0,9%	0,8%	0,8%	0,7%	0,0%	0,0%	0,0%	0,0%
$\Delta R4_{\text{cp.}}$	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	-0,9%	0,0%	-0,7%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%

Таблица Г.6 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "4-34-ГП-Н-Ми" к законопроекту "Об образовании"

Таблица Г.7 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "1-11-ОД-Н-Ми" к законопроекту "Об образовании"

Квота для класса	Бюджет времени, Td/Td_{\max}																			
	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
1-11-ОД	1	1	6	14	16	17	21	21	22	17	22	22	22	22	18	19	19	22	22	22
1-11-ОД-Н-Ми	1	1	6	12	14	14	18	18	19	14	19	19	19	19	15	16	16	19	19	19
Полнота отбора высказываний																				
R1	0,081081	0,081081	0,486486	0,675676	0,756757	0,756757	0,945946	0,945946	1	0,756757	1	1	1	1	0,864865	0,891892	0,891892	1	1	1
R2	0,123711	0,123711	0,43299	0,659794	0,783505	0,783505	0,979381	0,979381	1	0,783505	1	1	1	1	0,886598	0,938144	0,938144	1	1	1
R3	0,130435	0,130435	0,652174	0,73913	0,782609	0,782609	0,913043	0,913043	1	0,782609	1	1	1	1	0,869565	0,869565	0,869565	1	1	1
R4	0,324324	0,324324	0,756757	0,810811	0,837838	0,837838	0,945946	0,945946	1	0,837838	1	1	1	1	0,891892	0,891892	0,891892	1	1	1
R1 _{исх}	0,054054	0,054054	0,216216	0,72973	0,783784	0,783784	0,972973	0,972973	1	0,783784	1	1	1	1	0,810811	0,864865	0,864865	1	1	1
R2 _{исх}	0,020619	0,020619	0,14433	0,659794	0,731959	0,731959	0,989691	0,989691	1	0,731959	1	1	1	1	0,742268	0,855567	0,855567	1	1	1
R3 _{исх}	0,086957	0,086957	0,347826	0,826087	0,826087	0,826087	0,956522	0,956522	1	0,826087	1	1	1	1	0,869565	0,869565	0,869565	1	1	1
R4 _{исх}	0,054054	0,054054	0,378378	0,891892	0,891892	0,891892	0,972973	0,972973	1	0,891892	1	1	1	1	0,918919	0,918919	0,918919	1	1	1
Прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1$	8,1%	0,0%	40,5%	18,9%	8,1%	0,0%	18,9%	0,0%	5,4%	-24,3%	24,3%	0,0%	0,0%	0,0%	-13,5%	2,7%	0,0%	10,8%	0,0%	0,0%
$\Delta R2$	12,4%	0,0%	30,9%	22,7%	12,4%	0,0%	19,6%	0,0%	2,1%	-21,6%	21,6%	0,0%	0,0%	0,0%	-11,3%	5,2%	0,0%	6,2%	0,0%	0,0%
$\Delta R3$	13,0%	0,0%	52,2%	8,7%	4,3%	0,0%	13,0%	0,0%	8,7%	-21,7%	21,7%	0,0%	0,0%	0,0%	-13,0%	0,0%	0,0%	13,0%	0,0%	0,0%
$\Delta R4$	32,4%	0,0%	43,2%	5,4%	2,7%	0,0%	10,8%	0,0%	5,4%	-16,2%	16,2%	0,0%	0,0%	0,0%	-10,8%	0,0%	0,0%	10,8%	0,0%	0,0%
$\Delta R1_{исх}$	5,4%	0,0%	16,2%	51,4%	5,4%	0,0%	18,9%	0,0%	2,7%	-21,6%	21,6%	0,0%	0,0%	0,0%	-18,9%	5,4%	0,0%	13,5%	0,0%	0,0%
$\Delta R2_{исх}$	2,1%	0,0%	12,4%	51,5%	7,2%	0,0%	25,8%	0,0%	1,0%	-26,8%	26,8%	0,0%	0,0%	0,0%	-25,8%	11,3%	0,0%	14,4%	0,0%	0,0%
$\Delta R3_{исх}$	8,7%	0,0%	26,1%	47,8%	0,0%	0,0%	13,0%	0,0%	4,3%	-17,4%	17,4%	0,0%	0,0%	0,0%	-13,0%	0,0%	0,0%	13,0%	0,0%	0,0%
$\Delta R4_{исх}$	5,4%	0,0%	32,4%	51,4%	0,0%	0,0%	8,1%	0,0%	2,7%	-10,8%	10,8%	0,0%	0,0%	0,0%	-8,1%	0,0%	0,0%	8,1%	0,0%	0,0%
Средний прирост полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$(\Delta R1)_{cp}$	8,1%	4,1%	16,2%	16,9%	15,1%	12,6%	13,5%	11,8%	11,1%	7,6%	9,1%	8,3%	7,7%	7,1%	5,8%	5,6%	5,2%	5,6%	5,3%	5,0%
$(\Delta R2)_{cp}$	12,4%	6,2%	14,4%	16,5%	15,7%	13,1%	14,0%	12,2%	11,1%	7,8%	9,1%	8,3%	7,7%	7,1%	5,9%	5,9%	5,5%	5,6%	5,3%	5,0%
$(\Delta R3)_{cp}$	13,0%	6,5%	21,7%	18,5%	15,7%	13,0%	13,0%	11,4%	11,1%	7,8%	9,1%	8,3%	7,7%	7,1%	5,8%	5,4%	5,1%	5,6%	5,3%	5,0%
$(\Delta R4)_{cp}$	32,4%	16,2%	25,2%	20,3%	16,8%	14,0%	13,5%	11,8%	11,1%	8,4%	9,1%	8,3%	7,7%	7,1%	5,9%	5,6%	5,2%	5,6%	5,3%	5,0%
$(\Delta R1_{исх})_{cp}$	5,4%	2,7%	7,2%	18,2%	15,7%	13,1%	13,9%	12,2%	11,1%	7,8%	9,1%	8,3%	7,7%	7,1%	5,4%	5,4%	5,1%	5,6%	5,3%	5,0%
$(\Delta R2_{исх})_{cp}$	2,1%	1,0%	4,8%	16,5%	14,6%	12,2%	14,1%	12,4%	11,1%	7,3%	9,1%	8,3%	7,7%	7,1%	4,9%	5,3%	5,0%	5,6%	5,3%	5,0%
$(\Delta R3_{исх})_{cp}$	8,7%	4,3%	11,6%	20,7%	16,5%	13,8%	13,7%	12,0%	11,1%	8,3%	9,1%	8,3%	7,7%	7,1%	5,8%	5,4%	5,1%	5,6%	5,3%	5,0%
$(\Delta R4_{исх})_{cp}$	5,4%	2,7%	12,6%	22,3%	17,8%	14,9%	13,9%	12,2%	11,1%	8,9%	9,1%	8,3%	7,7%	7,1%	6,1%	5,7%	5,4%	5,6%	5,3%	5,0%
Абсолютный выигрыш в среднем приросте полноты при $\Delta t = 0,05 \cdot Td_{\max}$, %																				
$\Delta R1_{cp}$	2,7%	1,4%	9,0%	-1,4%	-0,5%	-0,5%	-0,4%	-0,3%	0,0%	-0,3%	0,0%	0,0%	0,0%	0,0%	0,4%	0,2%	0,2%	0,0%	0,0%	0,0%
$\Delta R2_{cp}$	10,3%	5,2%	9,6%	0,0%	1,0%	0,9%	-0,1%	-0,1%	0,0%	0,5%	0,0%	0,0%	0,0%	0,0%	1,0%	0,5%	0,5%	0,0%	0,0%	0,0%
$\Delta R3_{cp}$	4,3%	2,2%	10,1%	-2,2%	-0,9%	-0,7%	-0,6%	-0,5%	0,0%	-0,4%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%	0,0%
$\Delta R4_{cp}$	27,0%	13,5%	12,6%	-2,0%	-1,1%	-0,9%	-0,4%	-0,3%	0,0%	-0,5%	0,0%	0,0%	0,0%	0,0%	-0,2%	-0,2%	-0,2%	0,0%	0,0%	0,0%

Таблица Г.8 – Расчет среднего темпа прироста полноты отбора высказываний каждого вида с шагом $\Delta t = 0,05 \cdot Td_{\max}$ в классе "1-1-ОД-Н-Ми" к законопроекту "Об образовании"